Adaptation and habitat preference in a hybrid zone between *Bombina bombina* and *Bombina variegata* in Croatia

By

Catriona J. MacCallum

Thesis presented for the degree of Doctor of Philosophy
University of Edinburgh
1994

ABSTRACT

This thesis describes a hybrid zone between two taxa of toads, *Bombina bombina* and *Bombina variegata* in north eastern Croatia. The two taxa can be distinguished at four diagnostic enzyme loci. Clines at these loci are highly concordant; there is strong disequilibrium and substantial heterozygote deficit. Both the linkage disequilibrium and heterozygote deficit are asymmetric, being greater on the *bombina* side than on the *variegata* side.

Different habitats are identified across the zone and a strong association with the genotype of the populations sampled from them is found. This relationship is consistent across the hybrid zone. The cline is best described by a model which incorporates both a difference in gene frequency between habitats and a width which varies from place to place.

Mark recapture studies show extensive movement, which implies that the association between habitat and genotype is due to a habitat preference. Translocation experiments suggest that there is adaptation to the habitats. A habitat preference combined with mixing between habitats will inflate linkage disequilibrium over and above that expected from dispersal alone. Non-random mating and selection in relation to the environment will also contribute to the disequilibrium. As a result, inferences made using traditional cline models, where disequilibrium is mainly generated by dispersal, no longer apply.

These results are very different from those made from a previous analysis of the *Bombina* hybrid zone in Poland. There the cline showed a smoother transition of genotypes and a sharper step in gene frequency at the centre of the cline. The differences to the transect described here can be accounted for by a habitat preference.

A habitat preference has important implications for the mechanism of sympatric speciation since it will restrict gene flow between populations in different habitats.

Acknowledgements

I would like to thank the following people:

Prof. Nick Barton, my supervisor, for patiently guiding me through the maze of hybrid zone research and for providing many insights into other areas of population genetics. In particular, I would like to thank him for providing the 'analyse program' used to fit the cline.

Dr. Beate Nürnberger whose interest, encouragement and explanations have been a constant support to me throughout my Ph.D.Both Nick and Beate helped to collect all the field data in this thesis. I am extremely grateful to them for the input they provided both in Croatia and in Scotland; my thesis would be far thinner without them.

The work in Croatia could not have been carried out without Franjo Perović. He acted as our translator and our passport; gaining access for us into many areas and relieving us of many practical problems. His knowledge of natural history in general and specifically regarding Pešćenica were a constant source of information and delight and his field skills helped us in many ways. Both he and his wife Katica welcomed us unconditionally into their home and made us feel part of the family.

There are many people in Croatia I would like to thank. So many went out of their way to help us that it would be impossible to mention them all. In particular I would like to thank the following. Stjepan Tičerić, who provided us with a base for the van and our equipment when we were working in Perkoveć, Stejpan Dumbović who showed us various areas to work in, in the first year. Prof. dr. Sibila Jelaska who kindly provided us with space and equipment in her lab in Zagreb. Bojan Lazar and Eduarrd Kletečki from the Natural History museum in Zagreb helped us in the field. Finally I would like to thank the owners of the Domingo restaurant, Stejepan and Katica Rožić who provided many a welcome rest, always accompanied by good food and beer.

I am grateful to Dr Jacek Szymura for contributing all his enzyme data from 1979 to this analysis. I would also like to thank him for teaching me the electrophoretic and morphological techniques used here.

Dr Neil Sanderson helped a great deal in the first field season both collecting toads and showing us the area.

The Croatian museum of Natural History and the Croatian Ministry of the Environment were helpful in granting all the necessary permits.

Loeske Kruuk has provided new information regarding the hybrid zone at Pešćenica which she kindly allowed me to use. In particular she did the electrophoresis of the tadpoles which implied non-random mating.

Both Loeske and Beate went out of their way to help as I was finishing writing up. In particular I would like to thank Beate for providing me with the map of regions (Chapter 3) and Loeske for doing many last minute tasks.

Special thanks goes to Peter Jones who provided me with encouragement and support while I was writing up and took care of all the practical things I forgot about.

Finally, I would like to dedicate this thesis to my parents with thanks.

This work was supported by a Natural Environment Research Council studentship.

Contents

Tables and Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 1

Speciation, hybrid zones and *Bombina*

This thesis describes a hybrid zone between two species of toad, *Bombina bombina* and *Bombina variegata*. Hybrid zones are important as they provide 'windows of opportunity' for studying the evolutionary process (Harrison, 1990) and in particular are natural laboratories where the dynamics of gene flow between two populations can be examined (Barton and Hewitt, 1989; Hewitt, 1988). This introduction will first discuss the concepts and mechanisms of speciation and explain why hybrid zones form an integral part of their study. The observed characteristics and types of selection operating within hybrid zones will be outlined; special attention will be given to the role of environmental heterogeneity in determining and maintaining the structure of a hybrid zone. The inferences that can be made from a genetic analysis of hybridising populations, irrespective of the type of selection acting, will then be discussed. Finally there will be a review of the research already carried out on *Bombina*.

Speciation

Speciation is the process by which new species are generated. It has always been argued that in order to understand the process one first needs to define the concept. Unfortunately, however, the debate over an objective and universal definition of 'species', which has been running since the last century, has never reached a satisfactory conclusion. A brief perusal of the literature identifies many definitions and the arguments surrounding them (e.g. Carson, 1985; Cracraft, 1989; Dobzhansky, 1970; Mayr, 1942; Mayr, 1963; Paterson, 1985; Simpson, 1961; Templeton, 1989; White, 1978). In fact it has often been argued that there can be no rigid species definition at all, a view with which I agree. However the debate has been useful in clarifying issues which are of fundamental importance to evolutionary biology, namely how to identify and describe differences between populations, how those differences arose and how they are maintained. Species definitions are usually made so that specific questions can be answered whether from a taxonomic, ecological

or genetic basis. The most useful definition of species from a population genetics point of view, and the one that is most widely used by geneticists and ecologists alike, is that derived from the biological species concept (Mayr, 1942; 1963). Mayr defined species as "groups of actually or potentially interbreeding natural populations which are reproductively isolated from other such groups". It has been argued that the biological species concept (B.S.C.) is inappropriate in so many situations, especially for plants, and even for birds and mammals, that it should be abandoned (Carson, 1985; Cracraft, 1989). Despite its drawbacks the B.S.C. acts as a pointer with which to study the mechanisms of speciation. It defines species as populations with a common gene pool which cannot be shared with other biological species. This immediately poses questions as to how gene pools differentiate in the first place and how the barriers to gene flow are maintained.

The processes by which speciation are thought to occur can be viewed on two levels. Genetic mechanisms highlight how gene frequencies in a population might change sufficiently to allow for reproductive isolation. Geographic mechanisms provide alternative scenarios where the genetic processes can take place.

Genetic mechanisms of speciation

Genetic models of speciation have used the analogy of Wright's adaptive landscape to understand how speciation might occur (Wright, 1931; 1932; 1967). If species are seen as being in equilibrium on different fitness peaks then species formation is a problem of shifting from one peak to another. The change of a population from one peak to another will usually involve overcoming a selective barrier i.e. crossing the adaptive valley. Barton (1988) has reviewed four ways as to how this can occur; by random drift, changes in selection, through the application of quasi-neutral models and through the accumulation of incompatible mutations.

Drift is often thought of as the only way a population can cross an adaptive valley. As selection will always oppose a reduction of fitness, drift has to be strong enough to override its effect. As no natural population can be infinitely large, random sampling or drift will always play a role in altering gene frequencies. The smaller the population the greater the sampling effect and the easier it is to cross a selective barrier by chance. The problem is whether the effect of drift can be large enough for reproductive isolation to occur. Drift is often given as an explanation for the remarkable divergence between small isolated populations, whether this is spatial or temporal (Mayr, 1942).

The most often quoted example is that of the massive adaptive radiation of the Drosophilidae on the Hawaiian archipelago (Carson and Kaneshiro, 1976; Carson *et al.*, 1975; Templeton, 1980) where a combination of drift and other founder effects are proposed to provide opportunities for speciation. Whether drift alone can promote reproductive isolation, and hence speciation, is controversial (Provine, 1986; Barton and Charlesworth, 1984). Indeed, Barton (1988) concludes that there is no decisive evidence for it happening.

A population may reach a new adaptive peak directly through selection. If the selection pressures in an area change then the population can adapt through selection to those changes. This can have two consequences; 1) the total fitness of the population in relation to the equilibrium state it was in before, is reduced, effectively forcing it into a valley; it is then free to move uphill to a different peak or, 2) the relative fitness of a population increases due to a change in the selective environment. As with drift there is little evidence that selection causes speciation (rather than adaptation). The change in the population would have to be great enough to isolate it reproductively from populations still in the old state. Price *et al.* (1993) have provided a model whereby directional selection on one adaptive character causes correlated changes in other incidental characters despite the fact that changes in these alternative characters have a reduced fitness. This allows peak shifts in the correlated characters to occur. Given sufficient geological time this could result in the production of evolutionary novelties and promote the divergence of populations. Empirical evidence that changes in selection can cause peak shifts is rare. Barton (1988) cited two cases. Hybridisation between the sinistral and dextral forms of the snail *Partula suturalis* result in a lower reproductive output (Johnson, 1982; Johnson *et al.*, 1993 ; Murray and Clarke, 1980). Where the two forms meet it is hard to see how the rarer morph could be maintained in the population. The reason is provided by the presence of another species, *P. mooreana.* A mating between *P. mooreana* and the sinistral form of *P. suturalis* results in inviable offspring. When the frequency of *P. mooreana* is sufficiently high that the matings with the sinistral morph are more than a third then the dextral form of *P. suturalis* is at an advantage even when rare. Therefore, in the presence of *P. mooreana*, *P. suturalis* reaches a different equilibrium. Johnson *et al.* (1993) doubt, however, that differential coiling in this situation could result in speciation. The other example of peak shifts comes from Müllerian mimics (Turner, 1971). Here, the formation of new warning colours will be strongly selected against if it is not recognised by the predator. However new mimetic morphs may evolve if the models are sufficiently distasteful and/or common.

Quasi-neutral models allow populations to change state without having to cross an adaptive barrier. Here, change in a population occurs freely by moving around on the same fitness contour of the peak. Populations may have the same fitness in the same environment but this is due to different complexes of genes (Barton, 1986b). Crosses between the two populations may therefore be incompatible, resulting in a substantial reproductive barrier. Populations may diverge in a similar manner but due to sexual selection, where the 'runaway process' between female preference and male secondary sexual characters may occur at different rates in different populations. If populations are at different equilibrium levels of this process then pre-mating barriers will effectively isolate these populations when they reunite.

The accumulation of incompatible mutations between different populations may effectively isolate them (Müller, 1942; Wright, 1940). If populations are geographically isolated from each other it is not expected that changes in the genome will be identical even if their respective environments are. The changes may be sufficiently different that when the populations reunite the probability of restoring a common gene pool becomes unlikely. For example, the accumulation of incompatible mutations may account for Haldane's rule. This is the remarkable and consistent observation that where only one sex is sterile or inviable in the progeny of between-species crosses it is usually the heterogametic sex (Haldane, 1922). This has usually been attributed to dosage compensation, inactivation, or the higher mutation rate of the X-chromosome (Charlesworth *et al.*., 1987; Coyne, 1985; Coyne and Orr, 1989). However none of these theories offers a satisfactory explanation (Coyne, 1992; Coyne and Orr, 1989). Recently however it has been shown that Haldane's rule could be explained if alleles causing hybrid incompatibility behave as loss of function mutations in the foreign genetic background (Orr, 1994). Good examples where this may have happened come from hybridisations between *Drosophila* species (Orr, 1992; 1993).

Geographic mechanisms

Much discussion about speciation has involved the spatial conditions in which it is likely to happen. A randomly mating population can become reproductively isolated through the processes mentioned above. The prerequisite for these processes is the assumption that the population in question is isolated to some extent. However, a central issue is how much isolation is required for two populations to diverge. Can speciation occur when there is the homogenising effect of gene flow from another

source? The main discussion of speciation at this level rests on whether it can occur in allopatry, parapatry or sympatry.

Most speciation is thought to occur in allopatry. Here populations are separated by some physical barrier. Gene flow between the two populations ceases allowing each to diverge genetically via the mechanisms mentioned above. Mayr proposed that this was the only possible mode of speciation (Mayr, 1957). There is a large body of evidence that speciation does occur in allopatry. A good example is provided by the cichlid fishes in Lake Victoria (Greenwood, 1974). Populations of fish were isolated in a succession of changes in the water level which isolated distinct basins in the lake. Greenwood suggested that differentiation came about through drift, founder effects and local adaptation of each isolate. Reproductive isolation may have been caused by changes in courtship behaviour or male coloration, which differs between the many species. Differential predation is postulated as one cause of a change in male coloration. Males would tend to become less conspicuous in isolates where there was a high predation pressure (speciation via a change in selection). In areas where predation was not as intense an increase in colour would not result in a decrease of fitness and may instead result in increased fitness if there is a female preference for it. In this latter case one could envisage speciation between populations where there is little predation via the quasi-neutral model. There is good evidence in guppies (*Poecilia reticulata*) that differential predation does affect male coloration in this way (Endler, 1980; Endler, 1983). It has also been shown in laboratory experiments that female guppies prefer more brightly coloured males (Houde, 1988).

Parapatric speciation occurs when two populations are contiguous so that there is gene flow between them along a boundary. Divergence between populations happens in the absence of any geographic barrier. Parapatric speciation is the most relevant in the context of hybrid zones and will be dealt with subsequently.

Sympatric speciation occurs when the populations are mixed where theoretically any two individuals can meet; unlike parapatric speciation gene flow does not just occur along a boundary. Sympatric speciation amongst animals is notoriously contentious. Because of polyploidy it is much easier to envisage in plants where it is well understood (Stebbins, 1950). Models of sympatric speciation in animals usually invoke some form of disruptive or competitive selection. Maynard Smith (1966) was the first to show analytically that a stable polymorphism would result between two phenotypes controlled by a single gene if each genotype were fitter in a particular

ecological niche. The assumptions are severe; the numbers of individuals in each niche must remain constant and selection must be strong. However the polymorphism is more likely to arise if individuals choose the habitat to which they are best adapted. This has important implications for the research in this thesis and will be discussed more later (Chapters 4 and 6). There is some evidence for sympatric speciation, though often this can be explained using the model of allopatry (Futuyma and Mayer, 1980; Grant and Grant, 1989; Tauber and Tauber, 1989). Recently however it has been suggested that sympatric speciation may account for the divergence of cichlid species in two of the small volcanic crater lakes in Cameroon (Schliewen *et al.*, 1994). Molecular data on the species flock in each lake show they are monophyletic and most probably the result of one colonisation event. The physical and ecological nature of each lake provide no microgeographical barriers and neither could fluctuations in the water level produce separate basins (as in the example above). Given the high mobility of this species it is therefore unlikely that gene flow was initially impeded. Schliewen *et al.* (1994) proposed that the mechanism of speciation in this case was through ecological diversification but did not elaborate on how this resulted in reproductive isolation.

Hybrid zones

Hybrid zones and speciation

Hybrid zones are the result of gene flow between two contiguous differentiated populations. Smooth or steep clines can develop between the populations for a wealth of different phenotypic and genetic characters (Barton and Hewitt, 1985). They are evidence that populations can remain distinct despite gene flow. Hybrid zones form an intriguing phenomenon with regards to all the argument and debate surrounding speciation. By their very nature they defy the biological species concept, and yet the populations that are hybridising are often considered good species even by the proponents of the concept (Dobzhansky, 1940; Mayr, 1942; Wright, 1978). The nature of the origins of hybrid zones mirror the arguments about whether speciation can occur in parapatry as well as allopatry. Models to show that clines could develop in a continuous population state that population differentiation has evolved along an environmental gradient (Clarke, 1966; Endler, 1977). However most zones can be explained as regions of secondary contact, where divergence of the populations has initially occurred in allopatry (Barton and Hewitt, 1981c; White, 1978). It is difficult to distinguish between the two from the current pattern of variation (Endler, 1977).

The fate of hybrid zones depends on whether speciation can occur despite the homogenising effect of gene flow. In many cases it was thought that hybrid zones were areas which would promote the divergence of populations. As hybrids often show a lower reproductive output than the parental populations it was proposed that within hybrid zones there should be reinforcement of the mechanisms maintaining reproductive isolation (Dobzhansky, 1940). This has provoked a great deal of research, (Howard, 1993) and much controversy, (Barton and Hewitt, 1981c; Butlin, 1989; Moore, 1957; Paterson, 1978; Paterson, 1982). Hybrid zones may be stable for many thousands of generations without the development of pre-mating isolation and there is virtually no empirical evidence to demonstrate reinforcement exists (Butlin, 1989). Howard (1993) concluded that the reason reinforcement is still invoked is because it explains many of the patterns evolutionary biologists find in nature but he concedes that its future will be determined by difficult field and laboratory studies.

Hybrid zones do not only provide opportunities to study speciation. They are natural laboratories (Barton and Hewitt, 1989; Hewitt, 1988) where insights into the maintenance of genetic differentiation between populations can be gained. They provide a unique opportunity to study the genetic architecture between and within populations and the effect the dynamics of gene flow has on the populations concerned, in relation to both their genetic and their ecological environment.

Characteristics of hybrid zones

There has been some argument over the definition of a hybrid and what comprises a hybrid zone (Harrison, 1993). Barton and Hewitt (1985) define hybrid zones as narrow regions in which genetically distinct populations meet, mate and produce hybrids (Barton and Hewitt, 1985; 1989). Previously, hybrid zones were synonymous with areas of secondary contact (Mayr, 1942). More recently however, especially since the application of cline theory to hybrid zones, they are recognised by a series of concordant gradients of different characters i.e. clines. This is the definition I will use in this thesis.

Hybrid zones vary enormously. Some are described by changes in only a few characters. For example the different mimetic races of the butterfly *Heliconius erato* are distinguished at only six major genes affecting wing pattern (Sheppard *et al.*, 1985). Others, such as the hybridising taxa of *Bombina,* differ across a whole suite

of genetic, morphological and physiological characters (Szymura and Barton, 1986). However, despite the vast array of differences, many hybrid zones do share some common features (Barton and Hewitt 1985; 1989).

1. Most hybrid zones are narrow relative to the range of the species and to the distance moved in a generation. For example the alpine grasshopper, *Chorthippus parallelus parallelus* meets *C. p. erythropus* in a multiple zone along the Pyrenees. The taxa cover a wide area yet the cline for the X chromosome nucleolar organiser is 600m wide and the estimated dispersal rate per generation is 20-30m (Butlin and Hewitt, 1985; Hewitt *et al.,* 1988).

2. In hybrid zones where there are many differences between the populations, the change in characters tend to occur in the same place (i.e. the clines are coincident) and across the same distance (i.e. they are concordant). For example this is true for *Bombina* (Szymura and Barton, 1986; Szymura and Barton, 1991), and for the grasshoppers *Podisma pedestris* (Barton and Hewitt, 1981a). Hybrid zones where characters are not coincident have been the subject of particular interest. For example mitochondrial DNA is often not coincident with clines of nuclear genes as in the hybrid zone between *Mus musculus musculus* and *M. m domesticus* (Gyllensten and Wilson, 1987), or between red and Sika deer in Scotland (Abernethy 1994).

4. There are often strong associations (i.e. linkage disequilibria) between an array of quantitative and genetic traits. This is especially marked in the *Bombina* hybrid zones, between pairs of allozymes, between allozymes and morphological traits and also between other quantitative traits (Nürnberger *et al.,* 1994; Szymura and Barton, 1986; Szymura and Barton, 1991). Explanations for why these correlations are generated are discussed below.

Selection acting within hybrid zones

The outcome of hybridisation will depend on the interaction between different genotypes. These in turn will depend on the relative fitnesses of the populations and the recombinants between them (Hewitt, 1988; 1989). The following points outline possible fitness scenarios within hybrid zones.

1. The simplest scenario is when there is neutral diffusion. If the alleles distinguishing the races are equally fit in either genetic background then the parental

populations will gradually fuse and the clines between the populations will become shallower (Haldane, 1948; Nagylaki, 1975).

2. Alternatively, one allele may be fitter than the other, resulting in a wave of advance of one race and the possible extinction of the other (Fisher, 1937). This may help explain the spread of Sika deer, recently introduced into Scotland, over that of the existing red deer population (Abernethy, 1994).

3. There may be direct selection against heterozygotes or recombinants. Where hybrid unfitness is due to intrinsic genetic factors (as opposed to extrinsic ecological ones) then the resulting hybrid zone is defined as a tension zone. There are many examples of selection against hybrids within the literature ranging from hybrid mortality and sterility to varying degrees of hybrid inviability and morphological aberrations (Barton and Hewitt, 1989).

4. Frequency dependent selection is observed in some hybrid zones. The classic example is that of the Müllerian mimics of *Heliconius* butterflies (Mallet *et al.*, 1990; Mallet and Barton, 1989b; Sheppard *et al.*, 1985). *Heliconius* are unpalatable and the bright colours and patterns on the wings of these butterflies act as a warning to avian predators such as jacamars. Rare or unknown forms of patterning such as those generated by hybridisation between the mimetic races are unrecognised by the predators and selected against.

5. The parental genotypes may be fitter in different environments. In this case a hybrid zone will be at the transition of the two habitat types. Although a quarter of known hybrid zones are formed at environmental transitions (Hewitt, 1985, 1988), this does not necessarily imply adaptation to the different environments. This is discussed further below.

6. Hybrids may theoretically be fitter than either of the pure types in a narrow ecotone intermediate between the two parental environments. This is sometimes known as the 'geographically bounded hybrid superiority' model (Moore, 1977).

Mechanisms maintaining hybrid zones

Models of clines fall into two main categories; those that are dispersal independent and those that are dispersal dependent (Barton and Gale, 1993; Barton and Hewitt, 1985; Barton and Hewitt, 1989). Dispersal independent clines form where selection maintains a stable equilibrium at each locality. If this varies gradually then the cline will directly reflect local environmental conditions and be independent of how far individuals move. An example of a dispersal independent cline is Moore's model (1977) where hybrids are favoured in the intermediate habitat between the two parental types.

Models of dispersal dependent clines fall into two types. The first is typified by the model of Endler (1977) where clines are seen as a balance between dispersal and selection along environmental gradients or discontinuities. The second type of model is that of a tension zone where selection is against hybrids (Barton and Hewitt, 1981c, 1985, 1989; Bazykin, 1969; Key, 1968). In order for dispersal to have no effect on the maintenance of a cline its width has to be much greater than the characteristic scale of selection, l where $l \equiv \sigma/\sqrt{s}$; the dispersal rate, σ^2 is the variance in the distance moved within one generation i.e. the parent offspring distance while s is proportional to the strength in selection (Slatkin, 1973). It is therefore a measure which describes the distance over which selection can alter allele frequencies.

Tension zones

Barton and Hewitt (1985) believed that most hybrid zones were in fact tension zones. For a tension zone where selection is acting against heterozygotes the allele frequency, p, is described as

$$p=1/(1+\exp[-(x-x_0)/l) \qquad (1.1)$$

where $l = 4$ and is the characteristic scale of selection and x_0 is the arbitrary centre of the cline (Bazykin, 1969).

Tension zones, where clines are maintained by a balance between dispersal and selection against hybrids, have two distinctive features. They are not maintained by a response to local environmental conditions and so are able to move from place to place and also because they are free to move they will tend to do so in order to minimise

their length. In a survey of 150 different hybrid zones Barton and Hewitt (1985) showed that the majority of them could be explained as tension zones. (Barton and Hewitt, 1985, 1989; Barton and Gale 1993). Their reasons were as follows:-

1. There is evidence for dispersal.
Strong disequilibrium within many hybrid zones implies that most clines are not dispersal independent (Mallet *et al.*, 1990; Rand and Harrison, 1989; Szymura and Barton, 1986; 1991; Sites *et al.*, in press). As recombination would be expected to halve associations between loci each generation then the maintenance of disequilibrium is best explained through the dispersal of parental gene combinations into the zone (in conjunction with selection, see below).

2. There is evidence for selection against hybrids.
a. Hybrids are often inviable or sterile and many show developmental, morphological and genetic abnormalities (reviewed in Barton and Hewitt 1989). There are one or two exceptions to this, mainly in plants, where the hybrid populations are isolated from one or both parental types e.g. (Grant, 1971; Stebbins, 1950).
b. The fact that most clines are narrow relative to the range of the species implies that selection must be acting against hybrids. Given the historical age of many clines and the estimated dispersal rate of the species, then if there was neutral diffusion of the characters differentiating the populations the clines should be much wider than they often are. There are exceptions but often these are thought to involve underestimates of the dispersal distance for the taxa. This is notoriously difficult to measure empirically.

3. There is evidence that selection is against hybrids and not in relation to the environment.
a. Clines between races, whether genetic or phenotypic, are both coincident and concordant. If clines formed in direct response to the environment then they would not be expected to change at the same place or in the same way.
b. Clines have similar width and shape across different transects. If clines were determined through selection in relation to the environment they would vary from place to place depending on local environmental conditions. A good example of the similarity between transects comes from studies of the alpine grasshopper *Podisma pedestris* (Barton and Hewitt, 1981a; Nichols and Hewitt, 1986). The two races are distinguished by a Robertsonian fusion and the shape of this cline follows a smooth

sigmoid curve, 500-900m wide wherever they meet. Another example is the two Polish transects of the *Bombina* hybrid zone (Szymura and Barton, 1991).

One of the distinguishing features of tension zones is that they are free to move. They may move in response to local environmental conditions but can often be trapped by local density troughs or barriers to dispersal. Movement of genes from high to low density areas will tend to push the tension zone into the area of lowest density (Barton, 1979a; Hewitt and Barton, 1980). Empirical evidence for this comes from *Podisma pedestris* where detailed data on the density distributions show that the zone is held for large sections by low density (Barton and Hewitt, 1981b; Nichols and Hewitt, 1986). The fact that tension zones have the potential to move distinguishes them from dispersal independent clines which are ecologically determined. In the study on *P. pedestris* a detailed vegetational analysis suggested the zone was not determined by the local ecology (Nichols, 1985; Nichols and Hewitt, 1988).

The role of environmental heterogeneity

Since Barton and Hewitt (1985) published their survey there has been some debate over the role of environmental heterogeneity in determining and maintaining the position and structure of hybrid zones; that is whether in fact most zones are indeed tension zones. Selection in a hybrid zone can act in many ways as outlined above. The crucial distinction is whether there is selection in relation to the external environment, i.e. differential adaptation or selection against hybrids. Moore and Price (1993) have identified these two types of selection as exogenous and endogenous. They are typified by the models of Endler (1977) and Barton and Hewitt (1989) respectively. Both models are very similar except that for a given cline width they differ in the estimated strength of selection (May *et al.*, 1975) described by Moore and Price 1993). As errors in estimation of dispersal rates are far more likely to bias selection estimates the differences between the models are negligible. Although the different models may be used to estimate selection they cannot distinguish the nature of the selection acting. Evidence to identify how selection is operating will come from, for example, translocation experiments where hybrid and parental genotypes are transported to different parts of the zone and their relative fitness compared (Moore and Price 1993).

There is little evidence in the literature that selection is acting through adaptation. A quarter of known hybrid zones occur at environmental transitions. However it cannot

be presumed that this correlation implies extrinsic selection. There are a number of reasons why this association could be observed (Barton and Hewitt, 1985). Tension zones will move to the point where the two parental types are equally fit so that, for example, populations expanding after secondary contact may do so in parallel to environmental gradients. Hybrid zones are often found at local physical barriers (as in *Podisma*); here density troughs will explain the pattern (Hewitt, 1988). Also in many cases there is a broad environmental association but no close correlation of a character with a particular environmental gradient. Moreover if there were, one would expect the distribution of genotypes to show a more broken pattern reflecting the underlying environmental heterogeneity (Barton and Hewitt, 1985).

There are exceptions to this. The following brief survey highlights three studies where environmental heterogeneity plays an important part in the structure of a hybrid zone.

Mosaic hybrid zones

1. Harrison and Rand (1989; Rand and Harrison, 1989) provided evidence for one such case. Instead of the gradient models described above they have proposed a mosaic model of hybrid zones to explain the distribution of genotypes seen between the field crickets *Gryllus pennsylvanicus* and *G. firmus*. Although at low resolution there is a transition from one taxon to the other across the zone, a more detailed analysis revealed a patchy distribution which closely paralleled the soil type of the area. In general *G.firmus* was found on sandy soils while *G. pennsylvanicus* was more abundant on loam soils. Evidence from recapture data showed that 70% of marked adults were recaptured at least once,and often more, within 20m of the release point. Distances between paired neighbouring sites (where one site was *G.firmus*-like and the other was *G.pennsylvanicus*-like) ranged from 200m to 6km. This is close relative to the width of the cline. The correlation with soil type was consistent between all site pairs. Although they did not demonstrate that the habitat difference was adaptive it is hard to see why else there would be such a strong correlation.

Mosaic hybrid zones may approximate that of a smooth cline when the patch size is small relative to the dispersal distance. A mosaic model is more appropriate in the reverse situation. If patches are large enough compared to the dispersal distance then local gradients may accumulate at the patch boundaries which could be analysed using traditional cline methods. Harrison and Rand were unable to conclude whether clines

existed at the boundaries of their patches as there were uncertainties regarding dispersal distances and relative patch size.

This is not an isolated situation. Harrison and Rand provided examples of other mosaic hybrid zones (e.g. Gartside, 1980; Howard, 1986 among others). They conclude that a mosaic hybrid zones occurs "when closely related species that differ in habitat or resource utilisation patterns occupy a patchy environment". They believe that many such zones have gone unnoticed because the patterns of variation are not as striking as those formed by steep clines. Mosaic hybrid zones are important. It is recognised species are often not distributed smoothly but involve many independent encounters between local populations. Moreover they have important implications for the study of speciation itself as they may represent areas which increase the probability of reinforcement.

2. A hybrid zone between chromosome races of the lizard *Scleroporus grammicus* shares many features typical of other hybrid zones (Sites *et al.*, 1994). An initial analysis revealed steep concordant clines among three chromosme markers across a distance of about 2km with strong linkage disequilibrium between them (Sites *et al.*, 1993). Further analysis however showed that the zone was a mosaic of local patches (Sites *et al.*, 1994). When individuals were pooled across a patch size greater then 200m there was not only strong disequilibrium between the unlinked markers but there was also a significant deficit of heterozygotes and a highly significant association between karyotype and habitat such that one of the karyotypes was found more often on oak than expected. If patch size was less than 200m the heterozygote deficit and the habitat association disappeared though the disequilibrium remained. The effective selection against heterozygotes was estimated as 0.29 (this being defined as the total selection acting on all loci in linkage disequilibrium with the marker, see below). Sites *et al.* explained the strong selection both in terms of hybrid inviability (for which they had direct evidence) and also in relation to habitat. The fact that there was random mating and no habitat association within a patch size of less than 200m was explained by the short dispersal distance of these animals creating small panmictic neighbourhoods. Linkage disequilibrium within these patches is generated by immigration from other patches. The between patch divergence can be explained either by drift, with local clines forming at the boundaries, or in relation to habitat where the frequency of each habitat type within each patch varies. They conclude that, given the strong association of one of the metacentrics with oak, selection in relation to habitat rather than drift must have the greater effect.

This study is important on two accounts; first it demonstrates selection in relation to habitat; second it demonstrates a need to be careful when determining patch size in mosaic hybrid zones. The patch size was determined here by pooling successively over larger scales. If patches had been assigned from the start as smaller than 200m neither the habitat association nor the heterozygote deficit would have been revealed.

Associations between hybrid zones and ecological gradients

One of the lines of evidence demonstrating that most hybrid zones are tension zones comes from the fact that although many hybrid zones border a gross environmental transition, they do not closely map environmental gradients. An exception to this is provided by the detailed study of a hybrid zone between two types of woodpecker, the red and yellow shafted flicker (*Colaptes auratus*, Moore and Price, 1993). The two taxa differ in six plumage traits and several morphological traits. They are polymorphic at several allozyme loci, though these are not diagnostic. The variance of distance moved in a generation was estimated at 100km and estimates of gene flow are high. The zone extends for 4000km from Texas to southern Alaska. Moore and Price provided the following three lines of evidence that the structure of the hybrid zone is determined by 'exogenous' selection:-

1. Both mammalian and other avian range boundaries follow the course of the flicker hybrid zone. Assuming that the flicker hybrid zone is a result of secondary contact then this fulfils their prediction that other taxa should cluster around an ecotone if the populations diverged at the same time. Were the hybrid zone independent of the ecotone then one would not necessarily expect it to be in the same position.

2. The cline varies in width dramatically along its length. There is a correlation between the width and course of the hybrid zone and the steepness of several vegetational ecotones. The width is also correlated with a precipitation gradient.

3. The possibility that the hybrid zone follows a density trough is excluded despite the fact that it may decrease within the zone. The flicker hybrid zone bows substantially along its length, makes three major turns in orientation and varies dramatically in width. This therefore does not fulfil the predictions that a tension zone should move so as to minimise its length (in contrast with the *Podisma* hybrid zone mentioned above).

There are some reservations about their conclusions however. The fact that other avian and mammalian range boundaries coincide with that of the flicker hybrid zone does not imply that selection against hybrids is in relation to the ecotone. As mentioned before many hybrid zones meet at environmental transitions but reasons other than adaptation may explain it. For example the boundaries may also coincide due to a collective post glacial expansion.

Inferences from the shape of clines

Barton and Gale (1993) have demonstrated that the mechanism of selection in a hybrid zone has little effect on the shape of the clines. If selection is not too strong cline shape will not depend on the local population structure but can be approximated by diffusion (Nagylaki, 1975). The crucial parameter determining gene flow is the standard deviation of distance covered in one generation between parent and offspring. This allows inferences to be made regardless of the type of selection acting within the zone. This section will outline parameters important in describing a cline and what inferences can be made from the resulting cline shape.

Much of the theory describing clines has been based on a single locus model (Barton, 1979a; Bazykin, 1969; Endler, 1977; Haldane, 1948; Nagylaki, 1975). Clines using various models of single locus selection follow a smooth sigmoid curve. Smooth sigmoid clines appear as a straight line when plotted using a logit transformation ($z=\log_e(p/q)$, where p and q are allele frequencies). Therefore a cline maintained by selection against heterozygotes is linear with a slope $(\partial z/\partial x) = 4/w$ (where x is the distance along the cline and w is the cline width; Barton and Gale, 1993). However these models neglect the strong linkage disequilibria found between loci in many hybrid zones (e.g. *Heliconius* (Mallet and Barton, 1989a) or *Bombina* (Szymura and Barton, 1986; Szymura and Barton, 1991). These will have important consequences for the shape of the cline and the dynamics of gene flow. By investigating gene flow at linked loci Barton has developed an extensive multilocus theory which can be applied to many aspects of interest in hybrid zone research (Barton, 1979b; Barton, 1983). Linkage disequilibrium, the association between unlinked markers, is an important parameter in a genetic analysis of hybrid zones. In combination with cline shape it can be used to infer estimates of selection, the barrier to gene flow, rates of introgression and the number of genes under selection (Mallet and Barton, 1989a; Sites *et al.* 1994; Szymura and Barton, 1986; Szymura and Barton, 1991).

Apart from linkage disequilibrium the other measure that can be estimated directly from the pattern of genotypes across the zone is that of cline width. The width of the cline, in terms of allele frequencies, can be defined in two ways:

1. The distance over which gene frequencies change from some value $p=v$ to $p=(1-v)$, for example between the 20% and 80% points (Endler 1977). This measure can be used in practise but it is impossible to make explicit theoretical predictions from this (Barton and Gale, 1993).
2. When there is selection at a single locus the change in allele frequency will follow a smooth sigmoid curve. The width of the cline can be measured as the inverse of the maximum gradient of this cline (Barton 1989).

Dispersal rate

Associations between unlinked loci should halve every generation through recombination. The explanation for why associations between loci may remain is due to the continual influx of parental combinations of genotypes into the zone. It can intuitively be seen that the strength of disequilibrium in the centre of the zone must be proportional to the extent of the diffusion of parental gene combinations from either side of the zone and the rate of recombination within the zone. The relationship between disequilibrium, dispersal and recombination in a continuous population, where dispersal is approximated by diffusion in the centre of the zone is:-

$$D = \frac{\sigma^2}{r}\frac{\partial p}{\partial x}\frac{\partial u}{\partial x} \qquad (1.2)$$

(Barton, 1986a; Barton and Gale, 1993) where r is the rate of recombination between the two loci with allele frequencies p and u, σ^2 is the variance in parent offspring distance and $\frac{\partial p}{\partial x}\frac{\partial u}{\partial x}$ is the product of the slopes of the two loci in disequilibrium from one side of the cline to the other. These gradients are by definition the inverse of the cline widths so that:-

$$D = \frac{\sigma^2}{r w_p w_u} \qquad (1.3)$$

where w_p and w_u are the widths of the clines at these loci.

Barton and Gale (1993) have shown that there is good agreement between the theoretical prediction of linkage disequilibrium (under weak selection) and that observed when cline widths are simulated, whether selection is due to heterozygote disadvantage or epistasis. This means that dispersal rates can be estimated from observed values of disequilibrium no matter how the clines are maintained.

Empirical estimates of dispersal from mark recapture studies are notoriously difficult to measure as it is rare that individuals moving long distances will be picked up. Linkage disequilibrium can sometimes reveal information about long distance movement. For example in the zone between two hybridising species of fire-bellied toads in Poland (*Bombina bombina* and *Bombina variegata*) the edges of the cline for five diagnostic allozymes show a higher degree of linkage disequilibrium than expected. The most likely explanation for this is that a very small proportion of toads (about 1 in a thousand), move from one side of the zone to the other with their parental gene combinations intact, (Barton and Szymura, 1986).

Estimating the strength of selection

In a variety of models of selection the width of the cline is proportional to the ratio between dispersal and the square root of selection. The width of the cline itself is obtained directly from the inverse of the maximum slope (see above). Where the cline is maintained by a balance between dispersal and selection then the width is approximately equal to the characteristic scale, l, defined by Slatkin (1973). For heterozygote disadvantage:-

$$w = \sqrt{8\sigma^2/s} = 4l \qquad (1.4)$$

(Barton, 1979a; Bazykin, 1969; Szymura and Barton, 1986). Once an estimate of the dispersal rate is estimated from disequilibrium and the width is measured directly, then the strength of selection, at any one locus, maintaining the barrier to gene flow can be estimated. Although I have described only one type of selection here, Barton and Gale (1993) have shown that the relationship between dispersal, width and selection is robust for dispersal dependent clines maintained by other forms of selection.

Cline shape and barriers to gene flow

Selection on a single locus results in a smooth sigmoid curve (or straight line on a logit scale). However in many hybrid zones clines for both genetic and quantitative traits reveal a sharp step in the centre. This has been observed for example in *Bombina* (Szymura and Barton, 1986), *Ranidella* (Blackwell and Bull, 1978), *Urodema* (Baker, 1981; Barton, 1982),*Caledia* (Moran *et al.,* 1980; Shaw *et al.,* 1985) and *Podisma* (Jackson 1992). A sharp step simply indicates a barrier to gene flow. This could be caused in three ways; by a physical barrier (e.g. *Podisma*, reviewed by Jackson 1992), by long range migration, or through linkage disequilibrium. Linkage disequilibrium means that selection at one locus will cause parallel changes at other loci in disequilibrium with it. If selection is acting on more than one locus then the effective selection at any individual locus will be the net selection experienced at all other loci associated with it. This will inflate the observed selection seen at an individual locus even if that locus itself is neutral. This increased 'effective selection' will cause an increase in the rate of change in the centre of the cline, revealed as a sharp step (Barton, 1983; 1986a; Slatkin, 1975; Szymura and Barton, 1986). The more genes under selection, the greater the effective selection and the sharper the step.

Although most of the change in gene frequency will occur at the centre of the cline, foreign alleles may penetrate far into either side of the zone. The barrier causing the sharp change in gene frequency in the centre Δp is proportional to the gradient of change either side of the step, $\partial p / \partial x$, (Nagylaki, 1976). The strength of the barrier (B), can therefore be estimated as:-

$$B = \Delta p / (\partial p / \partial x) \qquad (1.5)$$

This can be interpreted as a distance and can be thought of as the length of unimpeded habitat that would present an equivalent barrier to a neutral allele (Barton, 1979b; Szymura and Barton, 1986, 1991). How much effect the barrier will have on gene flow across the cline will depend on the strength of the barrier and the dispersal rate. Barton (1979) showed that a neutral allele would be impeded by $(B/\sigma)^2$ generations. However any allele which is even slightly advantageous will hardly be delayed ($\approx \log[(B/\sigma)^2 \pi S/2\,]/2S$ generations).

Where selection is weak there is a relationship between the strength of the barrier to gene flow and the net selection maintaining the cline;

$$B = w(\overline{W}_{centre}/\overline{W}_{edge})^{1/r} \qquad (1.6)$$

where w is the width of the cline, $\overline{W}$ is the mean fitness of the population and r is the harmonic mean recombination rate between the marker and selected loci (Barton, 1986a; Barton and Bengtsson, 1986; Barton and Gale, 1993). Again this relationship does not depend on the type of selection acting; it applies to clines maintained by hybrid unfitness or by adaptation to the environment (Barton, 1986a). The relationship breaks down under strong selection at a single locus (approximately when selection is greater that 10% (Barton and Gale, 1993). However clines are often maintained by weak selection per locus and so this relationship can be used to find the mean fitness of a hybrid population. For example with *Bombina* the harmonic mean recombination rate was estimated as ≈0.25 and so given a cline width w, of 6.05km the mean fitness required to explain the observed barrier is 0.58 (Szymura and Barton, 1991). A distinction must be made here between the fitness of hybrids derived from this relationship and the estimate of selection inferred above from the dispersal rate. The selection inferred from the dispersal rate (i.e. from the observed disequilibrium) is a measure of the effective selection acting at a particular locus; selection here is estimated independently of disequilibrium (although it assumes the step is generated via disequilibrium) and is a measure of the relative fitness of the hybrid population. It is therefore an estimate of the net selection against hybrids.

Hybrid zones are complex phenomena. The populations are not reproductively isolated and yet can remain distinct. Although little information regarding their origins can be deduced from the current pattern of genetic variation, a great deal of information about how the populations remain differentiated can be inferred from the dynamics of gene flow across the zone. Many hybrid zones share common characteristics which can be used to determine the strength of the barrier to gene flow and the degree of difference which keeps them distinct. The differences between the pattern and distribution of genotypes in different hybrid zones may help reveal the nature of that barrier, whether through ecological adaptation or selection against recombinants.

The importance of adaptation in forming and maintaining distinctions between hybridising populations is the central question I wish to address in this thesis. I will

discuss investigations regarding the hybridisation between two taxa of discoglossid toads, *Bombina bombina* (L. 1761) and *Bombina variegata* (L. 1758) in Croatia. The next part of this introduction will briefly describe the study animal and review the extensive research already carried out on this zone.

The hybrid zone between *Bombina bombina* and *Bombina variegata* in central Europe.

The fire bellied toads *Bombina bombina* and *Bombina variegata* meet in a narrow hybrid zone that runs for 3-4000 km across eastern and central Europe. It extends from Austria along the southern edge of the Danube valley to the Black Sea and completely surrounds the Carpathian mountains along their foothills (Szymura, 1988). The two taxa show many morphological and ethological distinctions which appear to be adaptations to their different environments. In general *Bombina bombina* occupies lowland areas across eastern and central Europe and breeds in large permanent bodies of water whereas *Bombina variegata* lives in mountainous and hilly regions in the west and south, in temporary pools and small ponds. There is a large and growing literature on the *Bombina* hybrid zone. A thorough review is given by Szymura (1993).

Origins of the hybrid zone

The differences between the two taxa of *Bombina* in central Europe are thought to have risen in allopatry. Arntzen (1978) explained their divergence as a result of one or more Pleistocene glaciations (Fig. 1.1). At this time he suggested that *Bombina variegata* took refuge in the Appennine mountains and the northern Balkan mountains whereas *Bombina bombina* would have retreated either to the Hungarian Plains or the steppes bordering the Black and Caspian Seas. He concluded by suggesting that as the climate became more favourable there was a subsequent post-glacial expansion of *Bombina bombina* (at the expense of *Bombina variegata*), which colonised Bohemia and the northern side of the Bohemian-Moravian plateaus from the north, and the Hungarian Plains from the southern side of the Bohemian-Moravian

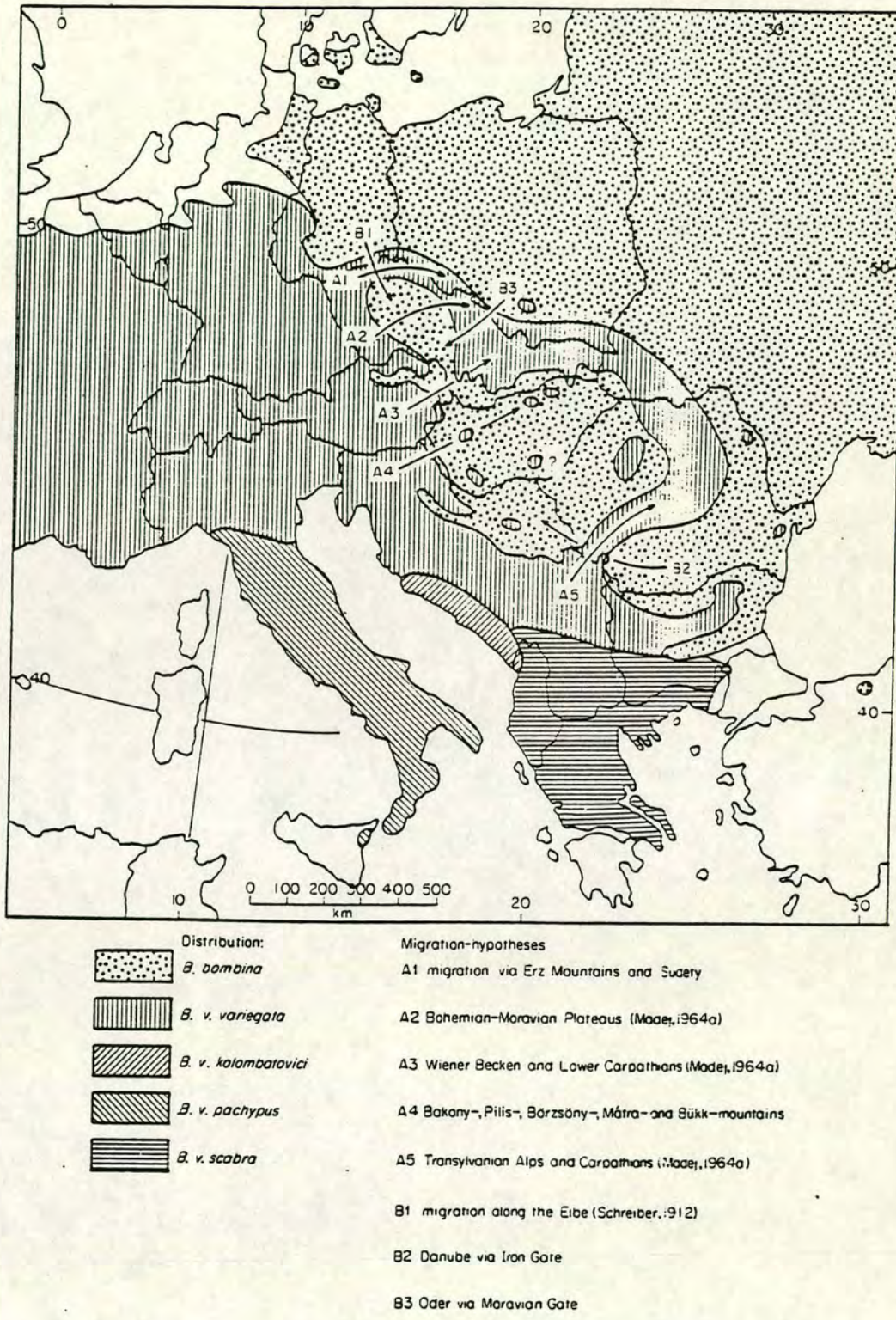


Fig 1.1 The distribution of the fire-bellied toad and the yellow bellies toad in central Europe. The arrows represent some hypotheses relating to their post glacial migrations (from Arntzen, 1978).

plateaus from the south-east. This would explain why there are populations of *Bombina variegata* present on isolated mountains today, for example the Bukk, Matre, Bakony and Mecsek mountains in Hungary, Fruska Gora in Bosnia, and the Bihor mountains in Romania. However, Maxson and Szymura (1984), using albumin as a molecular clock, estimated that the time the two species diverged was more likely to be during the Pliocene, within the last two million years. The subsequent Pleistocene glaciations would have affected only the distribution of the diverged *Bombina* species. Another analysis using electrophoretic comparisons of proteins provided further evidence of the divergence time of the two *Bombina* species but modified it to approximately 6.8 million years ago (Szymura, 1983). This biochemical evidence is supported in the fossil record where remnants of *B.bombina* like animals recovered in Poland and both *B. bombina* and *B. variegata* like animals from Czechoslovakia date back to the Upper Pliocene (reviewed by Szymura 1988; 1993). Although questions can be raised regarding the accuracy of both the biochemical and paleontological evidence Szymura concluded that the two *Bombina* species are older than previously assumed by both Mertens (1928) and Arntzen (1978).

Fossil records have also helped demonstrate how the Pleistocene glaciations have affected the distribution of *Bombina*. During this time the ranges of the toads contracted and expanded rapidly following the ice-sheet movements (Szymura 1988). Like Arntzen, Szymura suggested a refuge for *B. variegata* in the north western Balkans and one for *B. bombina* along the lower Danube and plains bordering the Black Sea. He assumed that *B. variegata* spread in two directions; one group migrated eastwards to the southern Carpathians and moved along them to the Moravian Gate and the eastern part of the Sudety Mountains, whilst the other group occupied western Europe and Italy north of the Po river. There are isolated populations of *variegata* scattered over the Danubian Plain today which are either related to the western type or are intermediate between the Carpathian and western types. These suggest that *variegata* once had a wider distribution in this region but has since been displaced by *bombina* as it invaded the lowlands of the Hungarian Plains. Today the ranges of the two species meet for thousands of kilometres across central Europe. When and where they first met is unknown but the taxa could well have been parapatric in south-eastern Europe towards the end of the Pleistocene. Contacts in the Danubian valley may therefore be several thousand years older than those on the slopes of the Carpathians (Szymura 1988).

The Pleistocene history and postglacial range expansion of *Bombina bombina* can account for differentiation between the *Bombina* groups present today. *B.bombina* and *B. variegata* are very distinct; the degree of divergence across 39 loci expressed as Nei's genetic distance = 0.49 (Nei, 1972; Szymura, 1988; 1993). Dendrograms using Nei's measure showed that the taxa could be further subdivided. *Bombina bombina* is a fairly uniform population, though a northern and southern group can be distinguished. *B. variegata* is more differentiated. It can be separated into four groups; the Carpathian, western, Balkan and Italian. The Carpathian and western groups are referred to as *B. v. variegata* while the Balkan and Italian groups are known as *B. v. scabra* and *B. v. pachypus* respectively. The Balkan and Italian groups are separated from the two northern groups. There is more divergence between the Balkan and Italian groups than there is between the western and Carpathian forms.

Differences between the taxa

It was C.L. Bonaparte in 1832 who first discussed the differences among the European *Bombina* (cited in Madej 1964), but a great deal of information regarding morphological, biochemical, ethological and genetic differences has been gathered since (Table 1.1). Hybridisation between the two species was first recognised with the discovery of morphological intermediates from the foothills of the Eastern Carpathian mountains (Horbulewicz, 1927) and other situations where the ranges of the two toads overlapped (Lac, 1961; Madej, 1965; 1964; Michalowski, 1961). This was finally confirmed by Szymura (1976) using allozyme electrophoresis.

Morphologically the most distinguishing feature between the two taxa is the colour and patterning differences on their bellies (Fig. 1.2). As their name suggests the yellow-bellied toad, *Bombina variegata* is identified by a large amount of yellow interconnected spots while the fire bellied toad, *Bombina bombina* is distinguished by smaller, red, discrete spots. Hybrids show an array of intermediate colours and patterns. The patterning is in fact so distinct that it acts like a fingerprint and can be used to distinguish between individuals in the field (Chapter 4). A score, based on the degree of interconnections between specific spots can be used to identify how hybrid an individual is (Michalowski and Madej 1969; Gollman 1984). This is a reliable indicator as it is highly concordant with the clines of diagnostic enzymes (Szymura and Barton, 1986; 1991) and other quantitative characters (Nürnberger *et al.*, 1994).

Measurement	*B. bombina*	*B. variegata*	Reference
Distribution	Lowlands of Eastern and Central Europe	Mountainous and hilly regions of western and southern Europe and the Carpathians	Andrin et al. (1984); Arntzen (1978)
Breeding sites	Large permanent waters	Temporary pools, small ponds	Madej (1973)
Habits	Largely aquatic	More terrestrial	Madej (1973)
Skin thickness (epidermis/dermis), μm	134.5 (22.8/111.7)	296.6 (65.2/231.4)	Czopkowa & Czopek (1955)
Breeding behavior	Prolonged breeder, territorial	Explosive breeder, non-territorial	Lörcher (1969); Szymura (unpublished)
Rate of calls (min^{-1})	22	95	Lörcher (1969)
Call duration (ms)	210	160	Lörcher (1969)
Fundamental frequency (Hz)	530	580	Lörcher (1969)
Sound pressure at 20 cm, 18°C (dyn cm^{-2})	10.5	1.8	Lörcher (1969)
Vocal sacs	Present	Absent	Boulenger (1886)
Lung volume in 4.5-cm toad (cm^3)	3.0–3.5	2.0–2.2	Lörcher (1969)
Mean fecundity (largest clutches observed)	363 eggs (509, 547, 689)	116 eggs (204, 233, 294)	Rafińska (1991) Rafińska (1991)
No. of eggs per clump (range)	32 (9–76)	17 (4–58)	Rafińska (1991)
Egg size (mm)	1.4	1.9	Rafińska (1991)
Development time at 20°C (egg to toadlet), days	73–75	61–63	Rafińska (1991)
DNA content per nucleus (pg)	18.8	21.1	Olmo et al. (1982)
Chromosome no. (identical karyotypes)	24	24	Morescalchi (1965); Wickbom (1949)
Nei's D, 29 loci	0.37–0.59		Szymura (1983, 1988)
Albumin distance (IDU)	2–4		Maxson & Szymura (1984)
mtDNA divergence (%)	5.6–7.		Szymura et al. (1985, unpublished)
No. of genes under selection	55 (26–88)		Szymura & Barton (1991)

Table 1.1 Differences between *B. bombina* and *B. variegata* (from Szymura 1993).

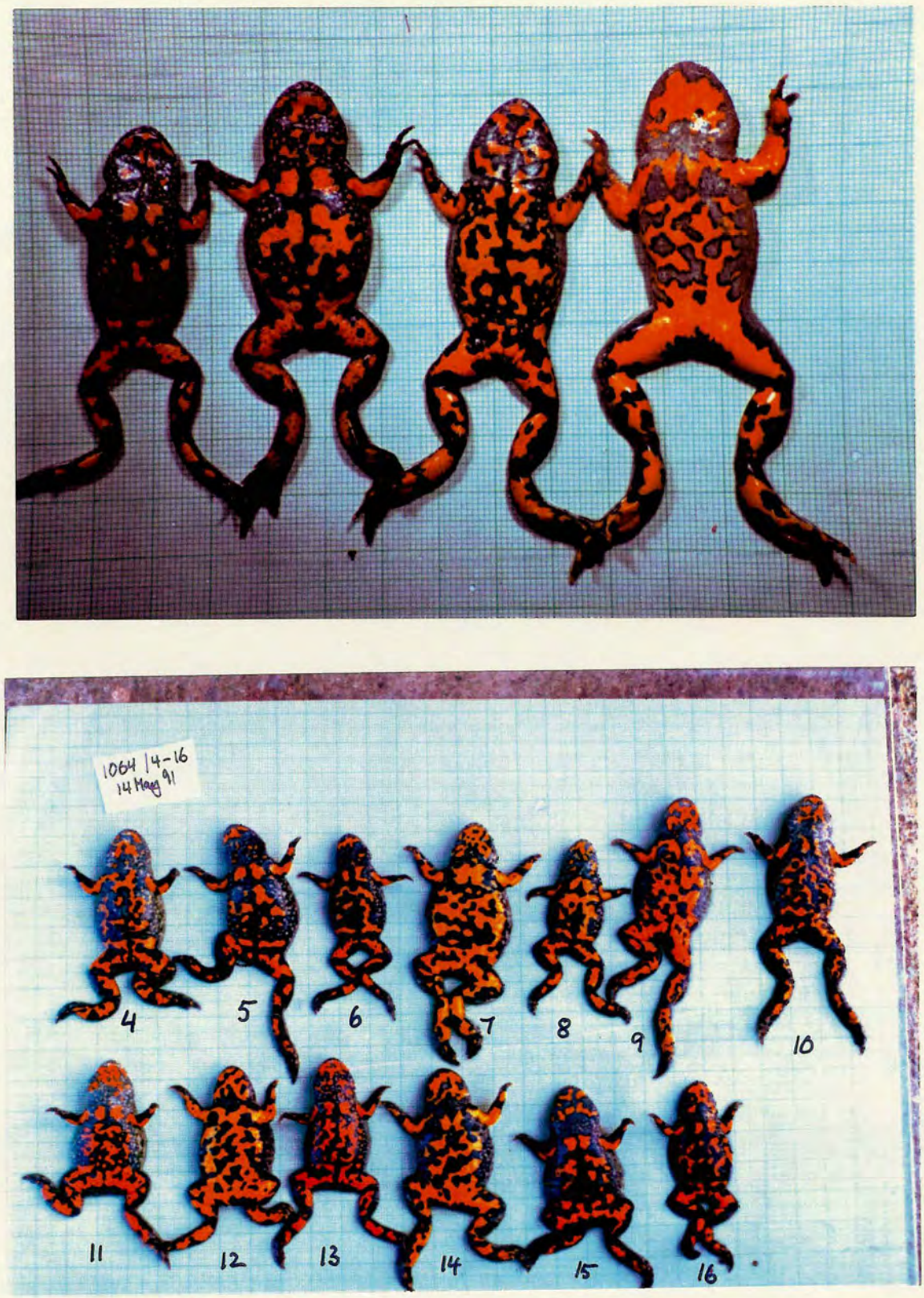


Fig 1.2 The differences in the belly markings of the fire-bellied toad *Bombina bombina* and the yellow-bellied toad, *Bombina variegata* and hybrid individuals. The top figure shows from left to right a typical *B. bombina*, two hybrids and a typical *B. variegata*. The bottom figure shows the range of belly markings from a hybrid population in the centre of the zone. The pattern is unique to each individual.

Many of the differences between the two taxa are thought to be adaptations to the habitats in which they live:-

1. *Bombina variegata* is generally found in upland areas and occupies small temporary pools and puddles, whereas *B.bombina* is found in more permanent water bodies in more lowland areas.

2. *B.variegata* lays larger eggs than *B. bombina* which reach metamorphosis more rapidly (Nürnberger *et al.,* 1994; Rafinska, 1991). This may well be an adaptation to the ephemeral nature of their puddle habitat (Seidel, 1982).

3. *B.variegata* have a thicker skin (Czopkowa and Czopek, 1955; Nürnberger *et al.,* 1994) which may reflect the greater risk of de-hydration. They also have a more vascularised lung than *B. bombina* again associated with their more terrestrial life style (Czopkowa and Czopek, 1955).

4. *B.variegata* have longer legs (Michalowski, 1961; Nürnberger *et al.,* 1994). A discriminant analysis of different skeletal proportions showed that *B. variegata* can be distinguished from *B.bombina* on the basis of the length of the tibiofibula and femur bones (Nürnberger *et al.* 1994).

5. The mating call varies between the two taxa. This reflects the fact that *B. bombina*, unlike *B. variegata* has distinct vocal sacs (Boulenger, 1886); cycle length, pulse duration and fundamental frequency have been shown to differ between taxa from transects in Poland (Sanderson *et al.,* 1992; Lorcher, 1969); *B.bombina* have longer calls of lower frequency with a longer duration between them. At the Pešćenica transect, in Croatia only cycle length differs between the taxa (Nürnberger *et al.,* 1994). It is difficult to surmise an adaptive reason for these differences but it is known that calls of lower frequency travel further and the presence of vocal sacs mean the calls of *B. bombina* are louder. More permanent bodies of water may well be further apart in which case *B.bombina* may have to attract other toads from a greater distance.

This array of putatively adaptive traits fit well with the known ecology and distribution of the taxa. This implies that the two taxa may well be fitter in their respective

habitats. If hybrids, which are intermediate for all the above traits, are adapted to neither habitat it may well be that selection acts along an environmental gradient.

The electrophoretic techniques developed by Szymura to distinguish between these taxa have provided a valuable tool to analyse the genetics of this hybrid zone (Szymura, 1976a; Szymura, 1976b; Szymura and Barton, 1986, 1991). Syzmura (1976) identified six enzymes which differed between *Bombina bombina* and *Bombina variegata* around Cracow, Poland. These enzymes involved a liver esterase (Est-B), Creatine Kinase (CK), Adenylate Kinase (AK), Lactate dehydrogenase (LDH), Malate dehydrogenase (MDH) and Glucosephosphate isomerase (PGI). Est-B was subsequently disregarded as the results it gave proved to be inconsistent (Szymura, 1976,b). Szymura demonstrated that the alleles at these loci were diagnostic. The centre of the zone had the highest proportions of heterozygotes. Although two toads were found to be heterozygous at all the loci examined the possibility that they were F1 hybrids was discounted as it was thought more likely that these were the result of numerous back crosses (more loci were examined than just the diagnostic ones). A subsequent analysis of linkage and inheritance of the five enzyme loci using both artificially created hybrids and individuals collected from the field revealed that both parental alleles were equally functional in F1 hybrids, that there was random assortment, and that inheritance followed classical Mendelian lines (Szymura and Farana, 1978).

Considering the vast array of differences between these taxa it is not surprising that they are often considered good species. However strictly speaking they do not conform to the biological species concept. Therefore throughout this thesis I will refer to them as taxa rather than species.

Genetic analysis of two transects in Poland

An extensive genetic analysis has been carried out at two transects in Poland, (Szymura and Barton, 1986; 1991). In the section above regarding inferences that can be made from clines I often cited *Bombina* as an example. Here I wish to outline in detail the inferences from these transects in order to compare them with results from the Pešćenica transect made in this thesis. The results from the Polish transects are given in Table 1.2. The main points are:-

	Estimates made in: 1991		1986	
Number of sites				
Cracow	49		29	
Przemyśl	30			
Number of samples				
Cracow	57		34	
Przemyśl	32			
Number of individuals				
Cracow	3,014		1,988	
Przemyśl	1,091			
Deviations from Hardy–Weinberg (F_{IS})				
Cracow	0.014	(−0.012–0.040)	0.017	(0–0.034)
Przemyśl	−0.01	(−0.040–0.040)		
Standardized gene frequency variance, estimated from discordance between loci (F_{ST})				
Cracow	0.0083	(0.0050–0.0119)	0.0067	(0.0034–0.0100)
Przemyśl	0.0239	(0.0164–0.0317)		
Dispersal rate, σ (km gen.$^{-1/2}$)	0.99	(0.82–0.114)	0.89	(0.79–0.94)
Cline width, w (km)	6.05	(5.56–6.54)	6.15	(5.45–6.45)
Barrier to flow into *bombina*, B_b (km)	260.2		160	(48–430)
Barrier to flow into *variegata*, B_v (km)	51.2	(22–81)	280	(48–400)
Harmonic mean recombination, r	0.250		0.123	
Mean fitness of hybrids, $\bar{W}_H/\bar{W}_P$	0.58	(0.54–0.68)	0.65	(0.60–0.77)
Effective selection, s^*	0.22	(0.15–0.29)	0.17	(0.16–0.18)
Number of genes under selection, n	55	(26–88)	300	(80–2,000)
Selection on selected loci, s	0.020	(0.014–0.030)	0.0027	(0.0005–0.0065)
Selection on marker loci, s_e	0.0037	(0.0015–0.0058)	0.0016	(0–0.0038)
Long-range migration, $\int m\,dx$ (km gen.$^{-1}$)	0.081	(0.040–0.300)		

Table 1.2 Results of the observed and inferred estimates from two Polish transects across the hybrid zone between *B. bombina* and *B. variegata* (from Szymura and Barton, 1991).

1. The clines at both transects showed a sharp step in allele frequency at the centre bordered by long tails of introgression. They were narrow and concordant and the widths of the clines at both transects were similar (6.05km in Przemyśl, 6.15km in Cracow).

2. All the clines reflected an environmental transition from habitats suitable for *bombina* in the south to those suitable for *variegata* in the north, but there was no direct correlation between the cline position and underlying environmental structure, which seemed to be uniform across the zone.

3. The estimated selection against hybrids was strong. The effective selection on the marker loci was 0.22 at Przemyśl and 0.17 at Cracow. The fitness of hybrid populations were estimated as 0.58 and 0.65 respectively.

4. There was direct evidence for selection against hybrids. Early embryonic mortality was increased in the hybrid zone. There was evidence for increased developmental and morphological abnormalities (reviewed in Szymura 1993).

5. There was random mating across the zone; populations within and either side of the hybrid zone were in Hardy-Weinberg equilibrium.

6. Linkage disequilibrium was strong; averaged across all pairs of loci, and standardised by gene frequencies, it reached a peak in the centre of the cline of 0.22 at Przemyśl and 0.17 at Cracow. Dispersal rates ($km.gen^{-1/2)}$ estimated from the disequilibrium values were 0.99 and 0.89 respectively.

7. A comparison with morphological data collected at the two transects 33 and 55 years ago showed that the clines have neither moved or widened (Szymura and Barton, 1991).

These transects are remarkable not only in the extensive information that can be gleaned from them but also in their similarity. Despite the ecological and possibly adaptive differences displayed by the two taxa it is reasonable to conclude that they are indeed tension zones. However not all transects of hybridisation between *Bombina* have revealed this type of pattern. The role of environmental heterogeneity between the taxa may be more important across some transects than others.

Comparison of Polish transects with others

Szymura has outlined three types of hybrid zones which occur in *Bombina*; smooth clines, mosaic zones and residual zones (Szymura, 1993). The first is typified by that of the Polish transects i.e. smooth narrow concordant clines. Szymura also analysed a transect at Pešćenica in Croatia (reviewed in Szymura 1993). The clines of allozymes were consistent with the pattern seen at the Polish transects except that the estimated width was wider (9.5km compared to 6.05km at Cracow and 6.15km at Przemyśl) and there was much more scatter around the cline than in Poland. He grouped this transect with those from Poland. However he proposed that the increased noise around the cline might be expected if the distribution of genotypes was in fact determined to some extent by the differences in habitat between the two taxa.

A mosaic distribution of genotypes occurs between the taxa in Slovakia. In the Slovak Karst area of Eastern Slovakia the hybrid zone contains pure individuals of both taxa despite the occurrence of extensive hybridisation (Gollmann, 1986). The explanation for this apparent anomaly may be related to the patchy environmental structure of this area which creates possibilities for habitat segregation among genetically differentiated demes. The resulting differing choices of breeding sites by individuals in relation to genotype could account for the sympatric co-existance of parental types in the presence of a majority of hybrids. A similar situation has recently been described in Kostajnica, at the border between Croatia and Bosnia (Szymura 1988; 1993). Here the centre of the zone contains populations with a very bimodal distribution. The populations deviate significantly from Hardy-Weinberg expectations. Szymura observed a strong association between habitat and genotype, and concluded that the mosaic distribution was related to habitat and limitation of breeding sites in time and space.

A residual zone could result from hybrid zones derived either from smooth clines or a mosaic pattern of genotypes. Two transects examined by Gollman (1984) are best explained in this way. In the Waldviertel in lower Austria and across a transect north of Vienna, geographically isolated populations of either *bombina*-like or *variegata*-like individuals exist which exhibit considerable morphological and genetic differences and yet may be as little as 1.7km apart. Populations were close to Hardy-Weinberg equilibrium. In no place were the parental types found together. Populations from each taxon are separated by farmed fields. The marginal populations bear traces of former hybridisation but not the transition of genotypes one would expect given the proximity of the pure types. This was related to a lack of suitable habitats due to

human intervention. It appears likely that prior to this time more habitats were available.

The differences between the transects suggest that the role of the underlying habitat structure in determining the position and structure between the hybrid zones of *Bombina* varies. A possible reason that the clines in Poland are best described as tension zones may be due to a smooth environmental gradient. Where this is perturbed a noisier distribution of genotypes is observed.

The *Bombina* hybrid zone provides a rare opportunity to study both the genetic dynamics between two taxa and the controversy over the balance of roles between selection against hybrids and selection in relationship to the environment. This thesis extends the analysis of the hybrid zone between *Bombina* across the transect in Pešćenica, Croatia. Chapter 2 describes the distribution and pattern of genotypes seen in this area. The roles of environmental heterogeneity, dispersal and adaptation are investigated in Chapters 3, 4 and 5. Inferences from these results and the importance of environmental adaptation in determining and maintaining the structure of hybrid zones, as well as implications for the mechanisms of speciation are discussed in Chapter 6.

Chapter 2

A hybrid zone between *Bombina bombina* and *Bombina variegata* in Croatia; the distribution and pattern of genotypes.

2.1 Introduction

This chapter describes the pattern of genotypes seen within the study site. The electrophoretic methods will be outlined and three genetic parameters will be looked at in detail to help explain the genotype distribution. These are concordance of allele frequencies across loci, deviations from Hardy-Weinberg proportions, and associations between loci (linkage disequilibria). Their importance is discussed elsewhere (Chapter 1). Each section will outline the methods and results separately. The parameters will initially be analysed in relation to the gene frequencies of the populations sampled, regardless of the geography of the sites. There will be a synopsis of results at this stage. The final section of the chapter describes how spatial information is included by fitting a cline in two dimensions. The genetic parameters mentioned above will then be assessed in the light of this new spatial information.

2.2 General methods and materials

2.2.1 The Pešćenica transect

The study site is based around a small village called Pešćenica (45° 36'N, 16° 10'E) about 20 Km south east of Zagreb in Croatia. It extends to Trebovec in the north (45° 43'N, 16° 19'E) and to the Kupa river valley in the south covering an area of 20 x 27km. Fig. 2.2.1 is a map of the general area. Unfortunately up to date maps are difficult to obtain. The ones describing this transect are British M.o.D. maps from 1940, though little has changed since. Most of the forest clearings and boundaries are still intact. The major difference was the creation of the Sava canal in 1972. This joined the Sava to the Odra river along most of its length.

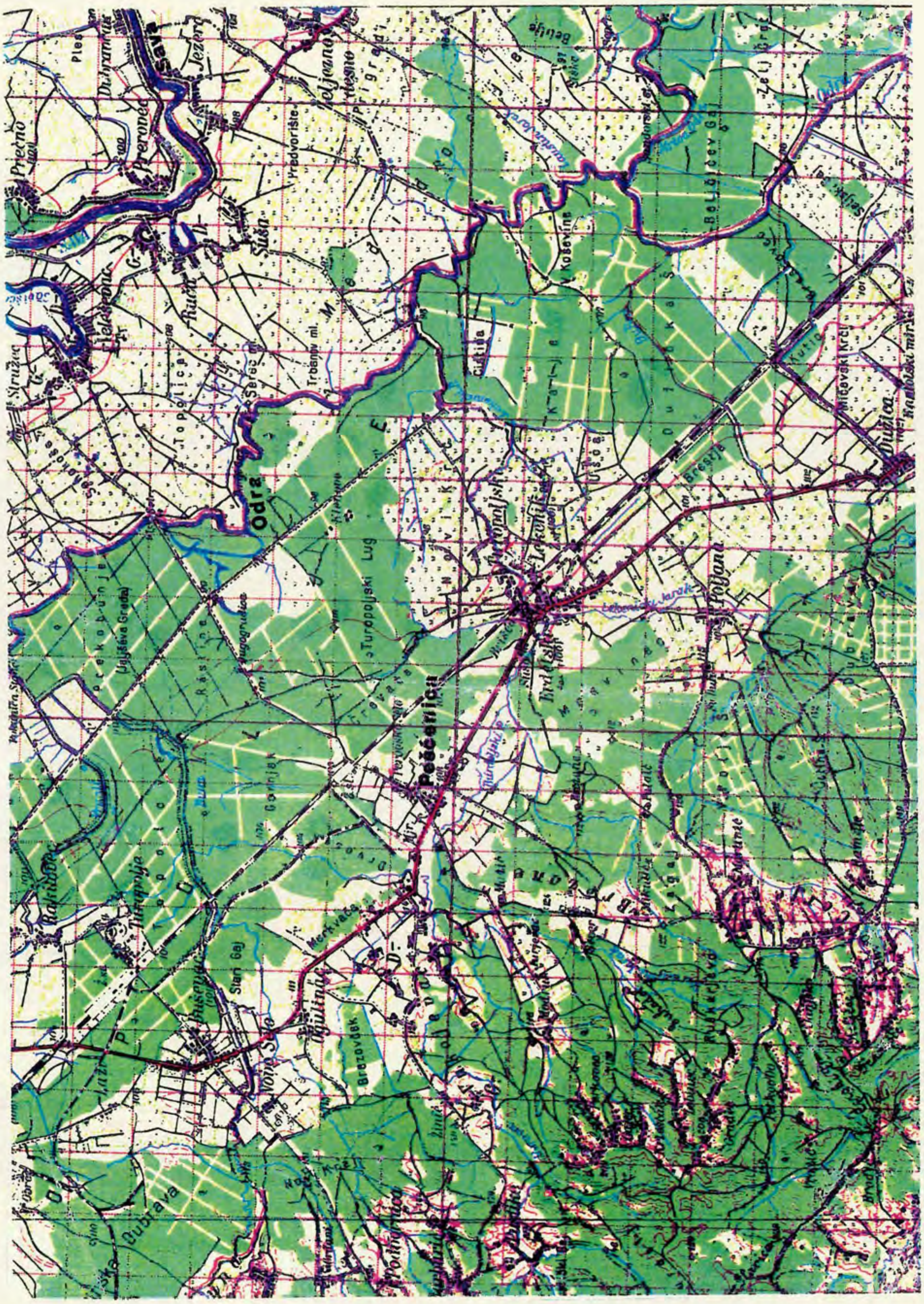


Fig. 2.2.1 Map of the study site at Peščenica. The overlay shows the variation in the mean allele frequency averaged across the diagnostic markers. Each pie is the proportion of *variegata* alleles in the population. Not all sites are shown; only those with a population size equal to or greater than five individuals. There are also some sites which extend beyond the range of this map (location of all sites given in Table 2.8.1). The arrows indicate nearby sites whose populations show a large difference in gene frequency.

2.2.2 Collecting and processing animals

The results presented in this chapter are from collections made over three seasons: 1979 (Szymura), 1991 (Barton, MacCallum, Nürnberger and Sanderson) and 1992 (Barton, MacCallum and Nürnberger). Szymura sampled twenty sites across the same area in 1979 (labelled 1-20). The results below show that samples do not differ significantly over the three years.

The field season extended from mid April to early July in 1991 and from May until July in 1992. This covers the beginning of the breeding season. Toads were collected from aquatic habitats. Sites varied from a collection of small puddles to discrete ponds but never covered an area greater than a hundred metres in radius. Individuals were brought back to the field laboratory for processing. Each individual was anaesthetised in 2% MS222 (3-aminobenzoic acid ethyl ester, Sigma). The anaesthetic lasted for approximately fifteen minutes; during this time their belly pattern was photographed and a toe was removed from either the left or right foot (left in 1991 right in 1992, from the toad's perspective). These were labelled and stored in liquid nitrogen for transfer to a -70°C freezer in Edinburgh. A number of morphometric measurements were also taken but will not be discussed in this study. The toads were kept in the field laboratory in plastic boxes with moist sponges for up to three days. They were then released back into the sites from which they were collected.

2.2.3 Site labelling system

Each site was given a name based on the order that the population was first sampled and the year it was sampled in. For example the first site in the first field season (1991) was named 1001, the second, 1002 and so on. The following year collections from these sites were labelled 2001 and 2002. New sites in the second year started at 2115 as 114 sites had been noted (though not necessarily collected from) the previous year.

Samples were collected from 147 sites (1764 individuals) in total. For all of the analysis in this chapter, apart from fitting the cline in two dimensions, estimates will be made for the 85 populations containing 5 or more individuals (1613 individuals in all).

2.2.4 Electrophoretic methods

Horizontal allozyme starch gel electrophoresis was carried out on the amputated toes collected in the 1991 and 1992 field seasons. Five loci were scored; adenylate kinase (Ak; EC 2.7.4.3), malate dehydogenase (Mdh-1; EC 1.1.1.37), lactate dehydogenase (Ldh-1; EC 1.1.1.27), isocitrate dehydrogenase (Idh-1; 1.1.1.42) and glucose phosophate isomerase (Gpi; EC 5.3.1.9). These were identified as diagnostic across the transects in Poland (Szymura 1976; a, b; Chapter 1). They are assumed to be neutral markers and it has been confirmed that they segregate independently (Szymura and Farana, 1978). Details of the protocols and recipes are given by Szymura (1976a; b; 1983; Szymura and Farana, 1978). The staining techniques are modified from those of Shaw and Prasad (Shaw & Prasad, 1970).

Szymura scored the toes collected in 1979 for the same loci apart from Idh. However he was also able to score creatine kinase (EC 2.7.3.2). I was unable to score this locus successfully. Szymura also scored some of the populations scored in 1991 and 1992 (Table 2.2.4.1). Szymura's data will be included in the analysis except where specified. The scores at each locus for each individual are given in Appendix 2.1.

2.2.5 Statistical methods

Most of the parameters in this chapter will be estimated using maximum likelihood (Edwards, 1972; Fisher, 1925; Hacking, 1965). The likelihood of a model is the probability of obtaining the data given that model. Comparisons between two models can be made directly from the ratio of their likelihoods, with the smaller likelihood as the denominator. It is both convenient and conventional to refer to the natural logarithm of the likelihood ratio, the log likelihood ($\Delta\log[L]$). A graph can be plotted of the log likelihood as a function of the parameters of interest. This function is known as the support curve. In likelihood terms the parameter value 2 units of log likelihood away from the maximum likelihood value is $e^2 = 7.4$ times less likely, if three units away then the hypothesis is $e^3 = 20$ times less likely, 5 units and it is $e^5 = 150$ times less likely. These values alone can be used to judge the relative plausibility of two hypotheses; alternatively they can be related to a standard significance test. For large samples the log likelihood is approximately distributed as $\frac{1}{2}\chi^2$ for the same degrees of freedom. Confidence limits with one degree of freedom around the

Table 2.2.4.1 Number of genes (N) scored and the frequency of *variegata* alleles (p), at each locus. Information is given on individual sites collected from in three different years. Pairs of sites in bold are the same site collected from in susequent years. $\bar{p}$ is the mean frequency averaged across all the diagnostic loci (Ak, Mdh, Ldh, Idh and Ck) pooled across years and weighted by sample size. Only one pair of sites differ in $\bar{p}$ between years (1/2054* see Table 2.3.3). Scorer indicates who carried out the electrophoresis; S is Szymura; M is MacCallum.

Site	Year	Scorer	Locus												
			Ak		Gpi		Mdh		Ldh		Idh		Ck		
			N	p	N	p	N	p	N	p	N	p	N	p	$\bar{p}$
1	1979	S	100	0.06	100	0.73	100	0.00	100	0.11	*	*	92	0.15	**0.079**
2	1979	S	96	0.08	98	0.80	98	0.00	98	0.03	*	*	98	0.15	**0.067**
3	1979	S	38	0.05	38	0.84	38	0.00	38	0.11	*	*	38	0.08	**0.059**
4	1979	S	44	0.07	44	0.93	44	0.00	44	0.09	*	*	44	0.11	**0.068**
5	1979	S	58	0.17	58	0.93	58	0.05	58	0.14	*	*	58	0.12	**0.121**
6	1979	S	70	0.14	70	0.83	70	0.03	70	0.13	*	*	70	0.19	**0.121**
7	1979	S	18	0.39	18	1.00	18	0.61	18	0.61	*	*	18	0.50	**0.528**
8	1979	S	20	0.50	20	1.00	20	0.95	20	0.90	*	*	20	0.85	**0.800**
9	1979	S	10	0.80	10	1.00	10	0.70	10	0.90	*	*	10	0.80	**0.800**
10	1979	S	14	0.50	14	1.00	14	0.21	14	0.14	*	*	14	0.29	**0.286**
11	1979	S	6	0.00	6	0.67	6	0.00	6	0.17	*	*	6	0.00	**0.042**
12	1979	S	10	0.80	10	0.90	10	0.50	10	0.70	*	*	10	0.60	**0.650**
13	1979	S	12	0.50	12	0.83	12	0.42	12	0.67	*	*	12	0.83	**0.604**
14	1979	S	24	0.67	24	0.83	24	0.58	24	0.67	*	*	22	0.41	**0.585**
15	1979	S	128	0.95	128	0.95	128	0.87	128	0.95	*	*	128	0.95	**0.932**
16	1979	S	32	0.84	32	1.00	32	0.69	32	0.91	*	*	32	0.69	**0.781**
17	1979	S	70	0.96	70	0.99	70	0.96	70	0.89	*	*	70	0.87	**0.918**
18	1979	S	50	0.90	50	1.00	50	0.88	50	0.90	*	*	50	0.86	**0.885**
19	1979	S	92	0.95	92	1.00	92	0.89	92	0.98	*	*	92	0.86	**0.919**
20	1979	S	28	0.96	28	0.96	28	0.96	28	1.00	*	*	28	0.89	**0.955**
1001	1991	M	56	0.79	56	0.95	56	0.71	56	0.70	*	*	*	*	**0.730**
2001	1992	M	14	0.50	14	0.93	14	0.79	14	0.64	6	0.83	*	*	
1002	1991	M	34	0.68	36	0.89	36	0.78	36	0.72	36	0.64	*	*	**0.684**
2002	1992	M	36	0.42	36	0.92	36	0.78	36	0.69	22	0.82	*	*	
1003	1991	M	30	0.60	20	1.00	30	0.67	30	0.77	28	0.82	*	*	**0.740**
2003	1992	M	44	0.61	48	0.98	48	0.79	48	0.79	38	0.84	*	*	

Table 2.2.4.1-continued

Site	Year	Scorer	Locus												
			Ak		Gpi		Mdh		Ldh		Idh		Ck		
			N	p	N	p	N	p	N	p	N	p	N	p	$\bar{p}$
1004	1991	M	6	0.50	6	0.67	6	0.33	6	0.50	4	0.50	*	*	**0.455**
1005	1991	M	2	0.00	2	1.00	2	0.00	2	0.50	2	0.00	*	*	**0.125**
1010	1991	M	2	0.00	2	1.00	2	0.00	2	1.00	2	0.00	*	*	**0.250**
1013	1991	M	10	0.10	10	0.90	10	0.00	10	0.00	10	0.00	*	*	**0.181**
2013	1992	M	8	0.37	8	1.00	8	0.37	8	0.25	8	0.50	*	*	
1014	1991	M	22	0.05	22	1.00	22	0.05	22	0.00	22	0.00	*	*	**0.023**
1015	1991	M	14	0.86	14	1.00	14	0.71	14	0.86	6	0.83	*	*	
1016	1991	M	6	0.17	6	1.00	6	0.00	6	0.00	6	0.00	*	*	**0.042**
1018	1991	M	2	1.00	2	1.00	2	0.00	2	1.00	2	0.00	*	*	**0.500**
1019	1991	M/S	10	0.20	6	0.83	10	0.20	10	0.20	10	0.20	4	0.50	**0.227**
1025	1991	M	2	1.00	*	*	2	1.00	2	1.00	2	1.00	*	*	**1.000**
1028	1991	S	12	0.92	*	*	12	1.00	12	1.00	12	1.00	12	1.00	**0.983**
1029	1991	M/S	74	0.96	46	1.00	74	0.97	74	1.00	72	1.00	28	1.00	**0.984**
1032	1991	S	2	0.00	*	*	2	0.00	2	0.00	2	0.00	2	0.00	**0.000**
1033	1991	S	4	0.00	*	*	4	0.00	4	0.00	4	0.00	4	0.00	**0.069**
2033	1992	M	26	0.08	26	0.88	26	0.00	26	0.19	18	0.06	*	*	
1035	1991	M	68	0.16	68	0.84	68	0.06	68	0.06	46	0.02	*	*	**0.080**
1036	1991	M	4	0.00	4	0.75	4	0.00	4	0.25	4	0.00	*	*	**0.062**
1037	1991	S	4	0.00	*	*	4	0.00	4	0.00	4	0.00	4	0.25	**0.050**
1038	1991	S	6	0.50	*	*	6	0.33	6	0.33	6	0.33	6	0.33	**0.367**
1039	1991	M	110	0.06	110	0.91	110	0.01	110	0.14	86	0.05	*	*	**0.062**
2039	1992	M	10	0.10	10	0.80	10	0.00	10	0.00	8	0.00	*	*	
1040	1991	M	72	0.07	72	0.86	72	0.01	72	0.14	62	0.06	*	*	**0.068**
2040	1992	M	8	0.13	8	0.88	8	0.00	8	0.00	8	0.00	*	*	
1041	1991	S	2	0.50	*	*	2	1.00	2	1.00	2	0.50	2	0.00	**0.600**
1042	1991	M	20	0.05	20	0.85	20	0.10	20	0.05	12	0.00	*	*	**0.056**
1043	1991	M	48	0.10	46	0.80	48	0.06	48	0.25	38	0.05	*	*	**0.126**
2043	1992	M	6	0.17	6	0.83	6	0.00	6	0.00	6	0.50	*	*	
1044	1991	M	38	0.26	36	0.75	38	0.13	38	0.34	30	0.20	*	*	**0.236**
1045	1991	M	20	0.10	18	0.83	20	0.05	20	0.20	20	0.00	*	*	**0.087**

Table 2.2.4.1-continued

Site	Year	Scorer	Locus												
			Ak		Gpi		Mdh		Ldh		Idh		Ck		$\bar{p}$
			N	p	N	p	N	p	N	p	N	p	N	p	
1046	1991	M	6	0.83	4	1.00	6	0.33	6	0.33	6	0.33	*	*	**0.458**
1047	1991	M	4	0.50	*	*	4	0.50	4	0.25	4	0.50	*	*	**0.437**
1049	1991	S	10	0.50	*	*	10	0.80	10	0.60	10	0.70	10	0.60	**0.640**
1050	1991	S	36	0.08	*	*	36	0.00	36	0.03	36	0.06	36	0.03	**0.039**
1051	1991	S	4	0.25	*	*	4	0.00	4	0.00	4	0.00	4	0.50	**0.150**
1052	1991	M/S	74	0.09	46	0.87	76	0.00	76	0.14	68	0.00	24	0.08	**0.063**
1053	1991	M	34	0.06	34	0.88	34	0.00	34	0.06	34	0.03	*	*	**0.037**
1054	1991	M	56	0.61	30	0.90	56	0.52	56	0.64	44	0.61	*	*	**0.594**
2054	1992	M	6	0.00	6	0.83	6	0.17	6	0.00	6	0.33	*	*	**0.125**
1055	1991	M	38	0.16	38	0.95	38	0.16	38	0.16	32	0.25	*	*	**0.139**
2055	1992	M	14	0.07	14	1.00	14	0.00	14	0.07	14	0.00	*	*	
1056	1991	M	38	0.29	30	0.93	38	0.18	38	0.26	28	0.11	*	*	**0.218**
1057	1991	M	8	0.00	8	0.50	8	0.00	8	0.00	*	*	*	*	**0.000**
1058	1991	S	2	1.00	*	*	2	1.00	2	1.00	2	1.00	2	1.00	**1.000**
1059	1991	S	6	0.67	*	*	6	0.67	6	1.00	6	0.50	6	1.00	**0.767**
1060	1991	M	2	0.50	2	1.00	2	0.50	2	1.00	2	1.00	*	*	**0.750**
1061	1991	M	4	0.00	4	0.75	4	0.25	4	0.25	4	0.25	*	*	**0.187**
1063	1991	M	72	0.39	72	0.88	72	0.18	72	0.33	72	0.32	*	*	**0.301**
2063	1992	M	2	0.50	*	*	2	0.00	2	0.00	2	0.00	*	*	
1064	1991	M	66	0.30	66	0.86	66	0.17	66	0.24	64	0.30	*	*	**0.252**
1066	1991	M	16	0.19	2	1.00	16	0.06	16	0.19	16	0.44	*	*	**0.219**
1067	1991	M	6	0.50	6	1.00	6	0.50	6	0.67	6	0.67	*	*	**0.583**
1068	1991	M	2	0.00	2	1.00	2	0.50	2	1.00	2	1.00	*	*	**0.625**
1069	1991	M	10	0.20	*	*	10	0.00	10	0.40	10	0.00	*	*	**0.150**
1070	1991	M	18	0.83	18	1.00	18	0.89	18	0.78	18	0.78	*	*	**0.819**
1071	1991	M	10	0.80	10	0.90	10	0.90	10	1.00	10	0.90	*	*	**0.900**
1072	1991	M	6	1.00	6	1.00	6	0.83	6	1.00	6	0.67	*	*	**0.875**
1073	1991	M	2	0.50	2	1.00	2	1.00	2	1.00	2	1.00	*	*	**0.875**
1074	1991	M	10	0.80	10	1.00	10	0.80	10	0.90	10	1.00	*	*	**0.850**
2074	1992	M	90	0.76	90	0.99	90	0.87	90	0.92	70	0.84	*	*	

Table 2.2.4.1-continued

Site	Year	Scorer	Locus												
			Ak		Gpi		Mdh		Ldh		Idh		Ck		
			N	p	N	p	N	p	N	p	N	p	N	p	$\bar{p}$
1075	1991	M	4	0.50	4	1.00	4	1.00	4	1.00	4	1.00	*	*	**0.875**
1076	1991	M	4	1.00	4	1.00	4	0.75	4	1.00	4	1.00	*	*	**0.938**
1077	1991	M	6	0.83	6	0.33	6	1.00	6	0.83	*	*	*	*	**0.889**
1078	1991	M	8	0.75	8	1.00	8	0.75	8	0.62	8	1.00	*	*	**0.781**
1079	1991	M	4	0.50	4	0.50	4	1.00	4	1.00	*	*	*	*	**0.833**
1080	1991	M	2	0.50	2	1.00	2	1.00	2	0.50	2	1.00	*	*	**0.750**
1081	1991	M	8	0.13	*	*	8	0.13	8	0.13	8	0.13	*	*	**0.125**
1082	1991	M	4	0.50	*	*	4	0.50	4	0.75	4	0.25	*	*	**0.250**
2082	1992	M	4	0.00	4	1.00	4	0.00	4	0.00	4	0.00	*	*	
1083	1991	M	2	1.00	2	1.00	2	1.00	2	0.50	2	0.00	*	*	**0.625**
1084	1991	M	6	0.33	*	*	6	0.17	6	0.33	6	0.17	*	*	**0.250**
1085	1991	M	6	0.50	6	1.00	6	0.50	6	0.67	6	0.00	*	*	**0.417**
1086	1991	M	2	0.50	*	*	2	1.00	2	0.50	2	1.00	*	*	**0.750**
1087	1991	M	10	0.90	*	*	10	0.90	10	0.70	10	0.80	*	*	**0.825**
1089	1991	M	2	1.00	*	*	2	1.00	2	1.00	2	1.00	*	*	**1.000**
1091	1991	M	8	0.88	8	1.00	8	1.00	8	0.88	8	1.00	*	*	**0.938**
1092	1991	M	2	1.00	*	*	2	0.50	2	1.00	2	1.00	*	*	**0.875**
1097	1991	M	12	1.00	12	1.00	12	1.00	12	0.92	12	0.50	*	*	**0.854**
1098	1991	M	2	1.00	*	*	2	1.00	2	1.00	2	1.00	*	*	**1.000**
1099	1991	M	46	0.76	46	0.98	46	0.63	46	0.83	46	0.93	*	*	**0.792**
2099	1992	M	20	0.70	20	1.00	20	0.85	20	0.85	20	0.80	*	*	
1104	1991	M	26	0.23	26	0.88	26	0.15	26	0.27	26	0.15	*	*	**0.202**
1105	1991	M	6	0.00	6	0.83	6	0.17	6	0.33	6	0.17	*	*	**0.167**
1109	1991	M	2	0.00	2	0.50	2	0.00	2	0.00	2	0.00	*	*	**0.000**
1110	1991	M	26	0.46	26	1.00	26	0.46	26	0.73	26	0.65	*	*	**0.577**
1111	1991	M	4	1.00	4	1.00	4	1.00	4	1.00	4	1.00	*	*	**1.000**
1112	1991	M	12	0.83	12	1.00	12	0.58	12	1.00	12	1.00	*	*	**0.854**
1113	1991	M	12	0.83	12	0.92	12	0.50	12	0.58	10	0.50	*	*	**0.609**
2011	1992	M	42	0.29	42	0.93	42	0.50	42	0.38	42	0.36	*	*	**0.381**
2012	1992	M	4	0.75	8	0.88	8	0.50	8	0.37	8	0.37	*	*	**0.464**

Table 2.2.4.1-continued

Site	Year	Scorer	Locus												
			Ak		Gpi		Mdh		Ldh		Idh		Ck		
			N	p	N	p	N	p	N	p	N	p	N	p	$\bar{p}$
2100	1992	M	12	0.75	12	1.00	12	0.83	12	0.83	12	0.92	*	*	**0.833**
2103	1992	M	22	0.14	22	1.00	22	0.18	22	0.32	22	0.27	*	*	**0.227**
2115	1992	M	14	0.07	14	0.79	14	0.00	14	0.21	12	0.08	*	*	**0.093**
2116	1992	M	36	0.11	36	0.83	36	0.00	36	0.17	28	0.04	*	*	**0.081**
2117	1992	M	26	0.12	26	0.81	26	0.08	26	0.08	16	0.00	*	*	**0.074**
2118	1992	M	4	0.50	4	1.00	4	0.75	4	1.00	4	1.00	*	*	**0.813**
2119	1992	M	14	0.07	14	0.71	14	0.00	14	0.07	14	0.14	*	*	**0.071**
2120	1992	M	12	0.25	12	1.00	12	0.00	12	0.17	8	0.00	*	*	**0.114**
2121	1992	M	16	0.00	16	1.00	16	0.00	16	0.06	2	0.00	*	*	**0.020**
2122	1992	M	8	1.00	8	1.00	8	1.00	8	1.00	8	0.75	*	*	**0.938**
2124	1992	M	4	0.50	4	1.00	4	1.00	4	1.00	4	0.75	*	*	**0.813**
2126	1992	M	16	0.44	16	1.00	16	0.88	16	1.00	16	1.00	*	*	**0.828**
2127	1992	M	8	1.00	8	1.00	8	1.00	8	1.00	8	0.88	*	*	**0.969**
2132	1992	M	4	0.75	4	1.00	4	1.00	4	1.00	4	1.00	*	*	**0.938**
2133	1992	M	16	0.94	16	1.00	16	0.94	16	0.81	16	1.00	*	*	**0.922**
2134	1992	M	18	0.72	18	1.00	18	1.00	18	0.94	16	0.88	*	*	**0.886**
2135	1992	M	22	0.09	22	0.91	22	0.05	22	0.14	8	0.13	*	*	**0.095**
2136	1992	M	6	0.67	6	1.00	6	1.00	6	1.00	2	1.00	*	*	**0.900**
2138	1992	M	2	0.50	2	1.00	2	1.00	2	1.00	2	1.00	*	*	**0.875**
2140	1992	M	12	0.42	12	1.00	12	1.00	12	0.92	8	0.75	*	*	**0.773**
2141	1992	M	8	0.62	8	1.00	8	0.88	8	1.00	2	1.00	*	*	**0.846**
2142	1992	M	6	0.17	6	1.00	6	0.17	6	0.17	6	0.17	*	*	**0.167**
2143	1992	M	72	0.18	78	0.87	72	0.07	78	0.15	24	0.13	*	*	**0.134**
2144	1992	M	*	*	4	1.00	*	*	4	1.00	*	*	*	*	**1.000**
2145	1992	M	12	0.17	12	0.92	12	0.25	12	0.25	12	0.17	*	*	**0.208**
2146	1992	M	12	0.67	12	0.75	12	0.92	12	0.92	12	0.67	*	*	**0.792**
2147	1992	M	20	0.30	20	0.85	20	0.45	20	0.60	18	0.44	*	*	**0.449**
2148	1992	M	8	0.13	8	1.00	8	0.00	8	0.13	8	0.25	*	*	**0.125**
2149	1992	M	2	0.50	2	1.00	2	1.00	2	0.50	2	0.00	*	*	**0.500**
2150	1992	M	6	0.17	10	0.90	10	0.10	10	0.20	10	0.40	*	*	**0.222**

maximum value of a support curve are those lying less than 2 log likelihood units below this maximum since the probability of obtaining $\frac{1}{2}\chi^2_1 >= 4$ is approximately 5%. These criteria apply for large samples only. As many of the samples dealt with in this chapter are small this is therefore a rough approximation and it may be better to interpret results directly in terms of likelihood.

Details of the estimates will be given in the appropriate sections below. All the parameters are calculated using the program, 'analyse' written by N.H. Barton in Pascal to run on Macintosh computers.

2.3 Estimating gene frequencies

The mean frequency of *variegata* alleles was estimated at each locus for each site (Table 2.2.4.1). The minimum and maximum frequency are 0 and 1 for all loci apart from Gpi where the *variegata* allele is polymorphic at around 70%.

A number of individuals were scored more than once for the same loci. This provides a useful estimate of the number of times I mis-scored individuals (Table 2.3.1). Out of a total of 78 individuals the duplicate scores for Gpi were the same. 1/78 individuals were misscored at Ldh, 2/78 at Mdh, 3/78 at Idh and 7/78 individuals at AK (1.28, 2.6, 3.8 and 9.0% respectively). This gives an overall error rate of approximately 4% (13/312). Mis-scoring at Ak and Mdh occurred in both directions in subsequent runs i.e. homozygotes were scored as heterozygotes and visa versa. The lack of consistency in mis-scoring should not affect the results of the genetic parameters I estimate below. However, the two individuals mis-scored at Idh were both identified as heterozygotes after being scored as homozygotes originally. Idh heterozygotes are difficult to identify as they produce a very faint signal. The identification of these two individuals as heterozygotes occurred on a subsequent scoring when I had increased experience. This may mean that heterozygotes were consistently miss-scored as homozygotes at the outset of electrophoresis. Estimates of deviations from Hardy-Weinberg seem to confirm this (section 2.6). The implications are discussed in that section. Mis-scoring will affect estimates of Hardy-Weinberg proportions if heterozygotes are scored as homozygotes or visa versa. Estimates of gene frequency and disequilibria will not be affected unless the same error occurs for the same individual across all loci; this is not the case.

Table 2.3.1 Number of *variegata* alleles scored at each locus on either two or three subsequent occasions. The total number of individuals scored more than once is 78. Gpi scored the same 100% of the time, Ldh 98.7%, Mdh 97.4%, Idh 96.2% and Ak 91% of the time. Bold type indicates those individuals scored differently at one or more loci. * represents missing data. Overall the error rate is approximately 4%.

Site	Ind	Ak			Gpi			Mdh			Ldh			Idh		
		No. of times scored														
		1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd
1039	**80**	**0**	**0**		**0**	**2**		**0**	**0**		**0**	**1**		**0**	**0**	
1053	19	0	0		2	2		0	0		0	0		0	0	
1053	21	0	0		2	2		0	0		0	0		0	0	
1053	22	0	0		2	2		0	0		0	0		0	0	
1054	**04**	**0**	**0**		**1**	**1**		**1**	**0**		**2**	**2**		*	**0**	
1099	21	2	2		2	2		1	0		2	2		2	2	
1099	22	2	2		1	1		0	0		2	2		2	2	
1099	**23**	**2**	**1**		**2**	**2**		**2**	**2**		**2**	**2**		**2**	**2**	
2011	01	0	0		2	2		2	2		1	1		1	*	
2011	05	0	0		2	2		0	0		0	0		0	0	
2011	09	0	0		2	2		0	0		0	0		0	0	
2011	**12**	**1**	**1**		**2**	**2**		**2**	**1**		**1**	**1**		**1**	**1**	
2121	05	0	0		2	2		0	0		0	2		*	*	
2136	02	1	1		2	2		2	2		2	2		2	*	
2143	32	2	*		1	1		2	*		1	1		2	*	
2159	06	0	0		1	1		0	0		1	1		0	0	
2159	09	0	0		2	2		0	0		0	0		0	0	
2159	10	0	0		2	2		0	0		1	1		0	0	
1029	17	2	2		2	2		1	1		2	2		2	2	
1029	21	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
1029	**23**	**1**	**2**		**2**	**2**		**2**	**2**		**2**	**2**		**2**	**2**	
1029	27	2	2		2	2		2	2		2	2		2	2	
1029	28	2	2		2	2		2	2		2	2		2	2	
1029	30	2	2		2	2		2	2		2	2		2	2	
1029	36	2	2		2	2		2	2		2	2		2	2	
1029	38	2	2		2	2		1	1		2	2		2	2	
1029	39	2	2		2	2		2	2		2	2		2	2	
1040	15	0	0		2	2		0	0		0	0		*	0	
1040	29	0	0		2	2		0	0		0	0		0	0	
1040	39	0	0		2	2		0	0		0	0		0	0	
1040	41	**0**	**1**		2	2		0	0		0	0		0	0	
1040	58	0	0		2	2		0	0		0	0		0	0	
1040	59	0	0		2	2		0	0		0	0		0	0	
1063	01	1	1		2	2		0	0		0	0		0	0	
1063	02	0	0		2	2		0	0		0	0		0	0	
1063	03	0	0		2	2		0	0		0	0		0	0	
1063	04	2	2		2	2		0	0		1	1		0	0	
1063	09	0	0		2	2		1	1		2	2		1	1	
1063	10	1	1		1	1		1	1		1	1		2	2	
1063	11	0	0		2	2		0	0		0	0		0	0	
1063	12	2	2		2	2		1	1		0	0		1	1	
1063	16	**0**	**1**		2	2		0	0		0	0		0	0	
1063	17	0	0	0	2	2	2	1	1	1	1	1	1	0	0	0
1063	18	**0**	**1**	**1**	2	2	2	0	0	0	0	0	0	0	0	0
1063	20	**2**	**2**	**1**	2	2	2	0	0	0	1	1	1	**2**	**1**	**1**
1063	22	1	1	1	2	2	2	1	1	1	1	1	1	1	1	1
1063	23	0	0	0	2	2	2	0	0	0	0	0	0	0	0	0
1063	24	1	1		1	1		0	0		0	0		0	0	
1063	25	0	0		2	2		0	0		0	0		0	0	
1063	29	**0**	**1**		2	2		2	2		2	2		1	1	

Table 2.3.1 continued

Site	Ind	Ak			Gpi			Mdh			Ldh			Idh		
		No. of times scored														
		1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd
1063	31	2	2		2	2		0	0		1	1		0	0	
1063	32	2	2	2	2	2	2	1	1	1	1	1	1	**2**	**1**	**1**
1063	35	1	1		1	1		0	0		0	0		1	1	
1063	37	1	1		2	2		0	0		2	2		1	1	
1064	01	0	0		2	2		0	0		1	1		1	1	
1064	02	0	0		2	2		0	0		1	1		1	1	
1064	03	2	2		2	2		2	2		2	2		1	1	
1064	04	1	1		2	2		0	0		1	1		1	1	
1064	05	1	1		2	2		0	0		0	0		1	0	
1064	08	2	2		2	2		0	0		0	0		*	1	
1064	09	1	1		2	2		0	0		0	0		0	0	
1064	10	0	0		2	2		1	1		1	1		1	1	
1064	11	0	0		2	2		0	0		1	1		0	0	
1064	12	1	1		2	2		1	1		0	0		0	1	
1064	13	0	0		2	2		0	0		0	0		0	0	
1064	14	2	2		1	1		1	1		2	2		2	2	
1064	15	1	1		2	2		0	0		0	0		0	0	
1064	16	0	0		2	2		0	0		0	0		0	0	
1064	22	0	0		2	2		0	0		0	0		1	1	
1064	23	0	0		2	2		1	1		2	2		1	1	
1064	24	0	0		1	1		0	0		0	0		2	1	
1064	28	0	0		2	2		0	0		0	0		0	0	
1064	31	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0
1064	32	1	1		1	1		0	0		0	0		0	0	
1064	33	0	0		1	1		0	0		0	0		1	1	
1064	34	0	0		2	2		0	0		0	0		0	0	
1064	38	1	1		2	2		1	1		0	0		0	0	
1064	40	1	1		2	2		0	0		1	1		0	0	
1064	41	0	0		2	2		0	0		0	0		0	0	

Ak - 7/78 (≈9%) scored differently on a subsequent gel run.
four of these changes were 0/1, two were 2/1.

Mdh - 2/78 (2.6%) scored differently 1/0 and 2/1

Idh - 3/78 (≈ 3%) scored differently - both were 2/1 changes

Ldh - 1/78 (1.3%) scored differently 0/1

The mean *variegata* allele frequency can be estimated for each site by averaging the values across all the diagnostic loci (hereafter known as $\bar{p}$; Table 2.2.4.1). As Gpi is not diagnostic it is therefore excluded. This average is weighted by sample size as not all individuals were scored for all loci. Therefore for sites 1-20 *variegata* frequencies are averaged across Ak, Mdh, Ldh and Ck whereas for most other sites the average frequency is across Ak, Mdh, Ldh and Idh. The interchange of Ck and Idh can be justified by the strong concordance between the loci (Section 2.5). A few sites have both Ck and Idh scored and so the mean is weighted across five loci. Populations collected from the same sites in consecutive years have been lumped together and given the name of the first year site. Apart from one site (1054 and 2054 see Table 2.3.2) $\bar{p}$ values between these populations did not differ significantly. Site 1054 and 2054 were kept distinct throughout the analysis.

Sites sampled in 1991					**Same sites sampled in 1992**				
Site	N	Med	Min	Max	Site	N	Med	Min	Max
1001	28	0.75	0.13	1.00	2001	7	0.67	0.50	0.83
1002	18	0.75	0.13	1.00	2002	18	0.71	0.25	1.0
1003	15	0.75	0.00	1.00	2003	24	0.75	0.5	1.0
1013	5	0	0	0.13	2013	4	0.32	0	0.88
1033	2	0	0	0	2033	13	0	0	0.37
1039	55	0	0	0.37	2039	5	0	0	0.13
1040	36	0	0	0.37	2040	4	0	0	0.13
1043	24	0.13	0	0.5	2043	3	0.25	0	0.25
1054	**28**	**0.56**	**0**	**1**	**2054**	**3**	**0.13**	**0**	**0.25**
1055	2	0	0	1.0	2055	7	0	0	0.13
1063	36	0.37	0	0.75	2063	1	0.13	-	-
1074	5	0.88	0.75	1.0	2074	45	0.83	0.37	1.0
1082	2	0.50	0.37	0.62	2082	2	0	0	0
1099	22	0.88	0.13	1.0	2099	10	0.75	0.62	1.00
No significant difference between gene frequencies at sites apart from **site 1054/2054: Mann-Whitney U = 6.0; z (for large samples) = -2.43 p<0.015**									

Table 2.3.2 Comparison of gene frequencies of individuals sampled from the same site in different years. The mean *variegata* gene frequency (averaged across the diagnostic loci, $\bar{p}$) was calculated for each individual and the median (med), minimum (Min), and maximum (max) value of $\bar{p}$ is given across all individuals at each site in both years, 1991 or 1992. N is the number of individuals sampled at each site. Sites are not significantly different from each other apart from 1054 which has a significantly lower $\bar{p}$ in 1992.

In the subsequent analyses the sample size for each population is given as the total number of alleles scored across all loci. This is the number of individuals in the sample multiplied by the number of loci scored for each individual multiplied by 2 (the number of alleles at each locus).

2.4 The distribution of genotypes

There are fewer samples whose populations are of intermediate gene frequency than those with more extreme frequencies (Fig. 2.4.1). Out of a total of 85 populations with 5 or more individuals, only 3 sites have a mean *variegata* frequency of 0.3-0.5. However this does not necessarily imply that there are fewer sites in the geographic centre of the hybrid zone: this can only be determined once the position of the cline is known.

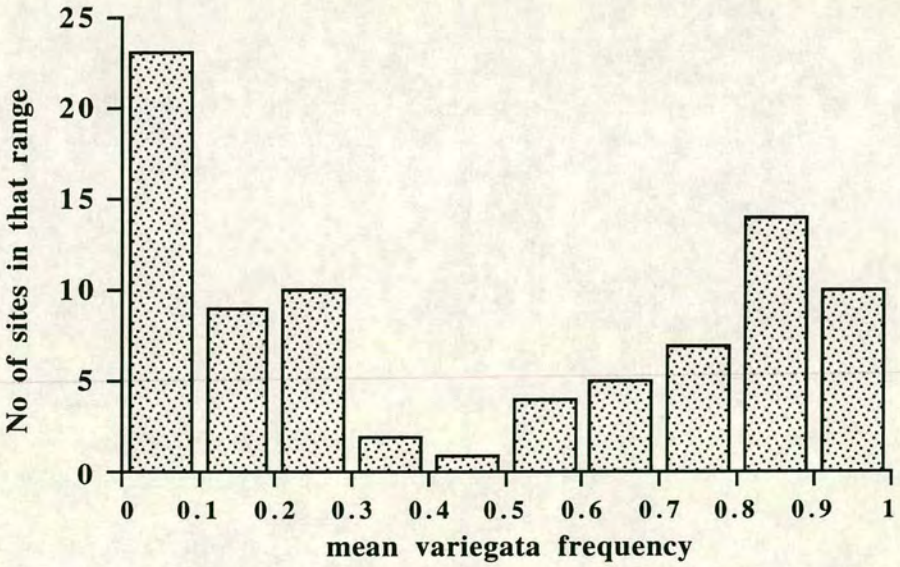


Fig. 2.4.1 The number of sites sampled over a range of gene frequencies. The mean *variegata* gene frequency is averaged across the diagnostic loci ($\bar{p}$). Fewer sites of intermediate gene frequency do not imply a dearth in the geographic centre of the zone.

The distribution of genotypes shows a cline that is broadly similar to the two *Bombina* transects in Poland (Szymura and Barton, 1986; 1991, Fig. 2.2.1, overlay). In general the more *variegata* -like populations are found in the upland regions to the south whereas the more *bombina*-like populations are found in the lowland areas to the north. However, unlike the transects in Poland, there are populations in the centre which show large differences in *variegata* frequencies despite being near each other geographically (see arrows on Fig. 2.2.1). This could be due to random fluctuations in gene frequency between neighbouring populations (though the differences are consistent across loci (Table 2.2.4.1); alternatively it may be due to some underlying environmental heterogeneity determining local gene frequencies. These ideas will be discussed and developed further towards the end of this chapter and in the following three chapters.

2.5 Concordance between loci

Apart from Gpi the *variegata* frequencies at each locus change in parallel across the zone. Their concordance can be demonstrated by plotting the mean frequency at each locus (p_i) against $\bar{p}$ (the mean *variegata* frequency averaged across all the diagnostic loci) of each population (Fig. 2.5.1). If these clines coincided exactly then all the points would lie along the diagonal. Despite a very close fit for each locus there is some scatter. There are two possible explanations for this variation; consistent differences in both the position and width of the clines at different loci or random fluctuations uncorrelated in space (Szymura and Barton, 1986; 1991). Both components can be described.

Differences in the shape of the clines can be estimated using a regression which describes the change in frequency at each locus as a function of the mean frequency:-

$$p_i = \bar{p} + 2\bar{p}\bar{q}\left[\alpha + \beta\left(\bar{p} - \bar{q}\right)\right]$$ (Szymura and Barton 1986, 1991). (2.5.1)

α denotes an increase in the *variegata* frequency of a sample resulting in a shift in the position of the cline in favour of*variegata*. The value of α is twice the shift in position measured in widths. For example, if α were 0.8 then the shift of that particular cline relative to the position of the average cline would be 0.4 widths, where the width is also that of the average cline. β reflects a change in the width of the cline

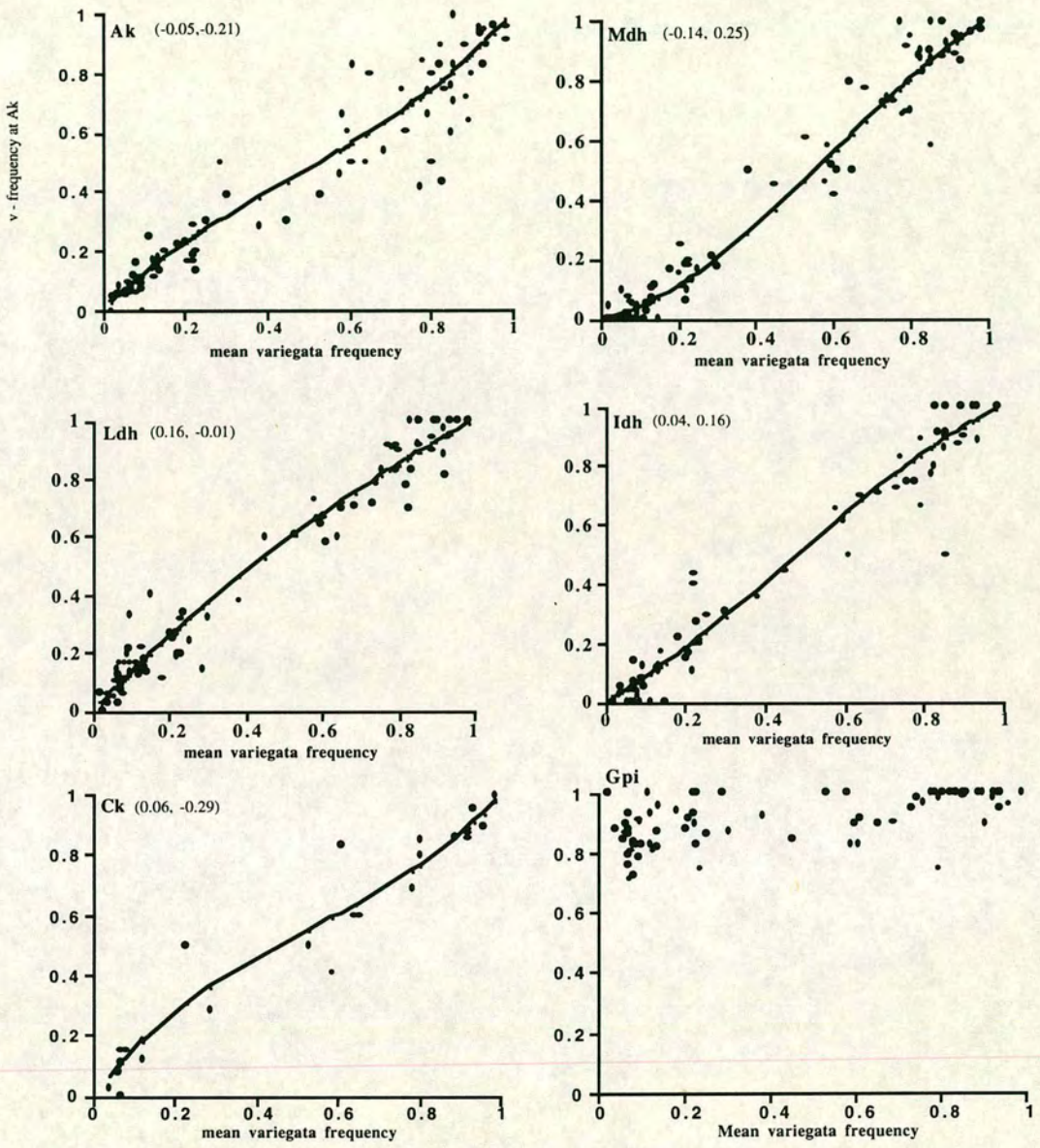


Fig. 2.5.1 Concordance across loci. The *variegata* allele frequency at each locus is plotted against the mean $\bar{p}$ (the average frequency across the diagnostic loci). The regression describes the change in frequency at the diagnostic loci as $p_i = \bar{p} + 2\bar{p}\bar{q}\left[\alpha + \beta\left(\bar{p} - \bar{q}\right)\right]$. The values of α and β are given in brackets on each graph (see section 2.5 Table 2.5.1 for details).

(defined as 1/maximum slope, Chapter 1). It is the amount by which the cline width is reduced, measured relative to the width of the average cline (Szymura and Barton 1986).

Despite allowing for different positions and widths of each cline there may still be some variation in allele frequency around this regression. These fluctuations can be described and estimated as the standardised variance F_{st}'. This is different from the usual estimate of F_{st}. If the loci fluctuated independently of each other then F_{st} would give var $(p)/\bar{p}\bar{q}$ where var (p) is the variance in allele frequency at an individual locus across sites. However as the change at each locus occurs in parallel the var(p) is the variance around the average cline shape (estimated from the regression for each locus) as p varies across the cline (Szymura and Barton, 1991, Barton and Gale 1993). This is a measure of the variance between loci but within sites.

Results

There are differences in the position and width of the clines at all the diagnostic loci. However they are small, emphasising the high degree of concordance between the diagnostic loci (Table 2.5.1; Fig. 2.5.1) . The largest change in position from the average is at Ldh which is shifted by 0.08 widths (α = 0.16) whereas the largest change in width is an increase of 29% (β = 0.29) at Ck. Over and above this regression there is a small amount of variation which is unaccounted for. The overall estimate of this, measured as F_{st}, is 0.00681. This differs significantly between loci across all sites (ΔL_{332} = 200.169).

Locus	α	β	F_{st}'
Ak	-0.05(-0.11,-0.00)	-0.21(-0.30,-0.12)	0.007
Mdh	-0.14(-0.18,-0.10)	0.25(0.18, 0.31)	0.013
Ldh	0.16(0.11, 0.21)	-0.01(-0.10, 0.07)	0.001
Idh	0.04(-0.02, 0.10)	0.16(0.06, 0.26)	-0.005
Ck	0.06(-0.06, 0.17)	-0.29(-0.47,-0.13)	-0.073

Total F_{st} = 0.00681 ΔL_{332} = -200.169

Table 2.5.1.Variation in the position(α) and width(β) of the clines at the different diagnostic loci. F_{st}' measures the variance of each locus around the average cline shape. Limits are given in brackets. (see Fig. 2.5.1 also)

The least variance is at Ck where $F_{st}' = -0.073$. This can be a negative measure as F_{st}' is estimated as the (observed variance) - (variance expected from sampling error). Therefore a negative value implies that the observed variance is less than expected. This is not surprising as there are far fewer populations scored for Ck than for the other loci (Appendix 2.1).The greatest variance is at Mdh where $F_{st}' = 0.013$.

2.6 Deviations from Hardy-Weinberg proportions

Method

Deviations from Hardy-Weinberg proportions were estimated for each population using Wright's measure, F_{IS} (Wright, 1922). F_{IS} is the fraction of heterozygotes in deficit or excess compared to the same population in Hardy-Weinberg proportions. For one locus with two alleles of frequency p and q the ratio of genotype frequencies is;

$$(p^2+pqF) : 2pq(1-F) : (q^2+pqF). \qquad (2.6.1)$$

F ranges from a minimum of either -p/q or -q/p to +1.

F_{IS} was estimated at each locus for every population containing more than five individuals using maximum likelihood. The support curve was generated by iterating the following formula for different values of F.

$$\text{Log } L = N_{pp}\log(p^2+pqF)+N_{pq}\text{Log}(2pq(1-F))+N_{qq}\log(q^2+pqF) \qquad (2.6.2)$$

Where p and q are the frequency of *variegata* and *bombina* alleles respectively and N_{pp}, N_{pq} and N_{qq} are the numbers of each genotype. Limits on the program required that F_{IS} be positive.

Results

F_{IS} pooled across all sites differs significantly between loci ($\Delta L_5 = 16.35$, $p<0.001$; Table 2.6.1). Populations are not literally lumped together as this would generate an inflated estimate of F_{IS} (Wahlund, 1928). Instead, F_{IS} estimates and limits are

Table 2.6.1 F_{IS} values and limits for all loci summed across all sites. Idh is significantly larger than the other loci.

Locus	F_{is} (limits)	ΔL
AK	0.069 (0.000-0.129)	3.21
GPI	0.056 (0.000-0.131)	1.33
MDH	0.145 (0.072-0.221)	8.61
LDH	0.074 (0.000-0.139)	2.91
IDH	0.294 (0.206-0.382)	25.86
Ck	0.000 (0.000-0.057)	-0.00

Test for heterogeneity among loci with Idh included $\Delta L_5 = 16.35$

Test for heterogeneity among loci without Idh, $\Delta L_4 = 4.70$

generated by summing the log likelihood values at each site. Ck shows no overall heterozygote deficit; Ak, Gpi and Ldh show similar, small levels (though Gpi is not significantly different from 0 [$\Delta L = 1.33$]) but both Mdh and Idh show large and highly significant values of F_{is}. F_{is} at Idh is especially large with a heterozygote deficit of approximately 30% ($\Delta L = 25.86$). If the enzyme markers were neutral and showed the same associations with linked loci they should have the same value of F_{is}. If F_{is} at Idh were the result of stronger selection then the cline would be expected to show a sharper transition across the zone than for other loci. The cline is indeed narrower than the average ($\beta = 0.16$; Table 2.5.1; section 2.5) but is not as narrow as Mdh ($\beta = 0.25$), which has the smaller heterozygote deficit. Both the clines for Idh and Mdh coincide with the clines at other loci (section 2.5) implying that the same order of selection is acting on each locus.

An alternative explanation accounting for the unusually high F_{is} at Idh is that this allozyme was mis-scored (see section 2.3). It is difficult to prove that I was mis-scoring heterozygotes as my error rates when rescoring individuals are similar to those for the other allozymes (Table 2.3.1). However there is some evidence for a consistency in mis-scoring. More importantly if the way F_{is} changes with gene frequency is different from the other loci then there would be reasonable grounds for excluding it from the main part of the analysis. If Idh is not included in the above analysis then there is no significant heterogeneity between loci ($\Delta L_4 = 4.7$).

Obtaining an overall estimate of F_{is} at each locus is not informative as it combines information across all sites. Any heterozygote deficit may be caused by assortative mating or selection. In each case populations either side of the zone (where $\bar{p}$ approaches 1 or 0) would be expected to approximate Hardy-Weinberg proportions whereas the central populations would be more likely to deviate. It is difficult to show how F_{is} changes across the zone in general when many of the sites have a small sample size. A clearer picture can be obtained by combining the F_{is} values of populations within a certain range of gene frequencies. This effectively increases the sample size within each frequency class. Populations were assigned to one of seven groups. The first contains populations whose $\bar{p}$ values range between 0 and 0.1. The remaining six groups are separated by a frequency range of 0.15.

F_{is} varies between loci at particular $\bar{p}'$ values ($\bar{p}'$ is the mean of the $\bar{p}$ values in each of the seven groups). For all loci except Idh F_{is} increases for those populations of intermediate gene frequency. Idh is the only locus where F_{is} decreases for populations of intermediate gene frequency (Table 2.6.2, Fig. 2.6.1). For this reason and those mentioned above Idh will be excluded from the subsequent analysis.

$\bar{p}'$	Ak	Gpi	Mdh
0.068	0.017 (0.000, 0.125)	0.042 (0.00, 0.413)	0.081 (0.000, 0.390)
0.160	0.167 (0.334, 0.310)*	0.113 (0.00, 0.288)	0.280 (0.100, 0.466)*
0.301	0.267 (0.063, 0.461)*	0.076 (0.00, 0.359)	0.315 (0.083, 0.534)*
0.487	0.411 (0.000, 0.770)	0.000 (0.00, 0.593)	0.130 (0.000, 0.551)
0.626	0.000 (0.000, 0.194)	0.111 (0.00, 0.4120)	0.166 (0.000, 0.359)
0.771	0.000 (0.000, 0.157)	0.000 (0.00, 0.465)	0.019 (0.000, 0.176)
0.914	0.032 (0.000, 0.160)	0.000 (0.00, 0.284)	0.094 (0.000, 0.259)
$\bar{p}'$	Ldh	Idh	Ck
0.068	0.069 (0.000, 0.189)	0.298 (0.092, 0.5255)*	0.000 (0.00, 0.204)
0.160	0.129 (0.000, 0.271)*	0.525 (0.339, 0.6871)*	0.000 (0.00, 0.165)
0.301	0.152 (0.000, 0.357)	0.201 (0.000, 0.4163)	0.300 (0.00, 0.867)
0.487	0.670 (0.243, 0.914)	0.100 (0.000, 0.6726)	0.000 (0.00, 0.521)
0.626	0.045 (0.000, 0.239)	0.2424(0.000, 0.4653)*	0.047 (0.00, 0.441)
0.771	0.013 (0.000, 0.182)	0.156 (0.000, 0.3627)	0.052 (0.00, 0.413)
0.914	0.000 (0.000, 0.126)	0.322 (0.000, 0.5952)*	0.000 (0.00, 0.085)

Table 2.6.2 F_{is} estimated by maximum likelihood at individual loci. Populations of a similar $\bar{p}$ are grouped together; $\bar{p}'$ is the mean of each group. Limits ($\Delta L = \pm 2$) are given in brackets. A * indicates F_{is} values which are significantly different from 0.

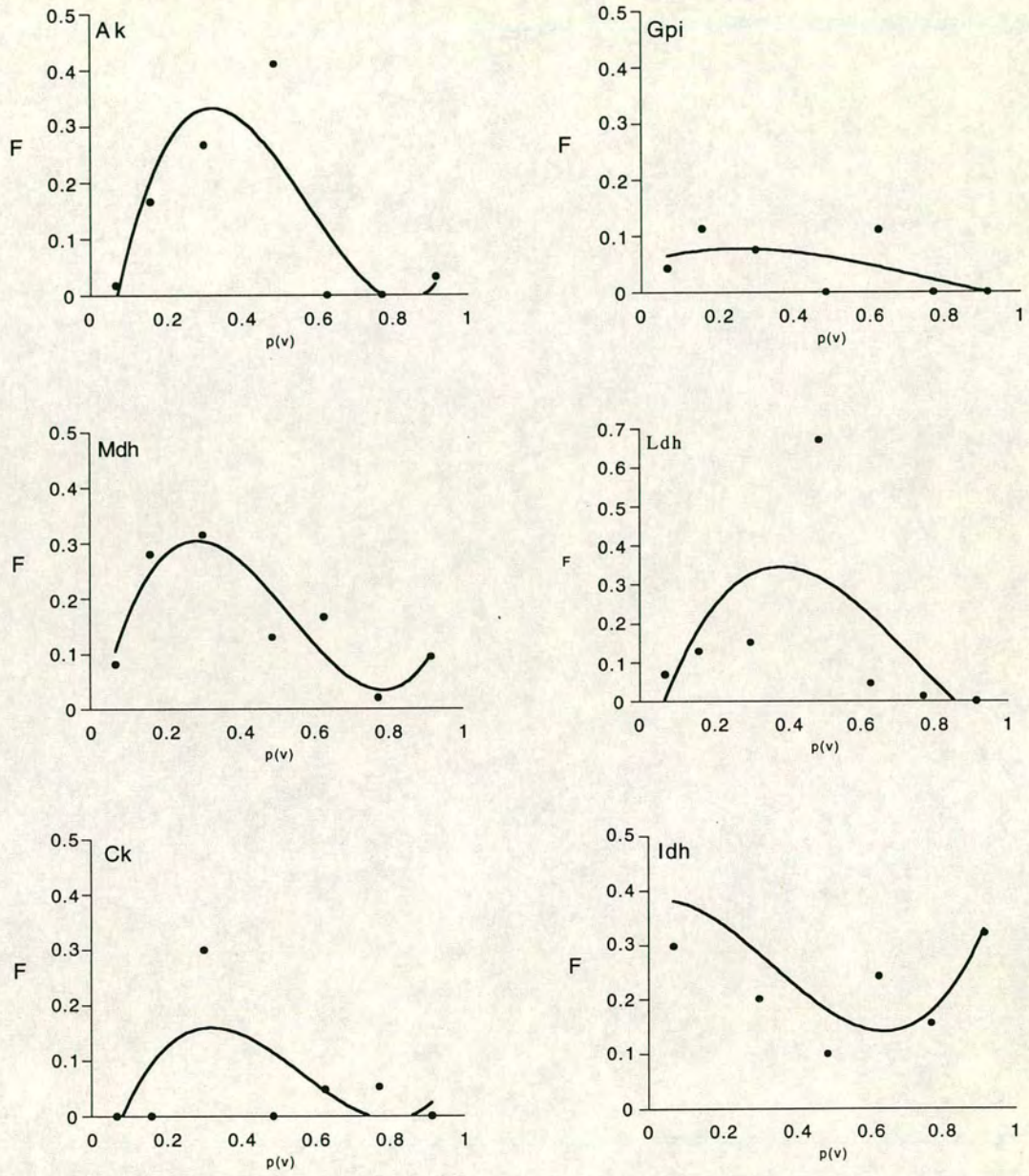


Fig. 2.6.1 How F_{is} at individual loci varies with $\bar{p}$'. Each data point represents a number of populations of similar $\bar{p}$ pooled together. A polynomial is used to illustrate the pattern at each locus. Only Idh shows a distinctly different shape (see section 2.6 for details).

It is difficult to see an overall pattern of how F_{is} changes with gene frequency by looking at the individual loci (Table 2.6.1). However F_{is} can be estimated across all loci by summing the log likelihoods of values at individual loci (Table 2.6.4, excluding Idh). Overall the total F_{is} does not vary significantly between sites (ΔL_{84} = 43.48), and it is significantly different from 0 at relatively few individual sites (Table 2.6.3). When the likelihoods of F_{is} are summed over all loci and across populations with a similar $\bar{p}$ (as above) then a distinct pattern emerges:-

Fis increases as $\bar{p}'$ approaches 0.5 i.e. towards the centre of the zone but is significantly different from 0 only for *bombina*-like populations (Fig. 2.6.2, Table 2.6.4). F_{is} does not differ significantly across loci or between sites within each group of populations. A cubic polynomial best describes the data; it explains more of the variation than a quadratic fit ($F_{1,4} = 5.16$; $p<0.1$) but is not significantly improved by a quartic fit ($F_{1,3} = 0.64$). There are a number of points to note:-

1. The heterozygote deficit is large at its peak. For populations with a mean *variegata* frequency of 0.33, there is a 26% reduction in heterozygotes.

2. The shape of the polynomial describing the change in Fis with gene frequency is asymmetric. There is a sharper reduction in heterozygotes for *bombina*-like populations i.e. for those populations of a lower gene frequency (0.00-0.33) that there is for *variegata* like populations i.e. those populations with higher gene frequencies (1-0.33).

The expression derived from this curve can be used to generate the expected Fis for any particular gene frequency (Fig. 2.6.2):-

$$F_{is} = 2.83\bar{p}^3 - 5.27\bar{p}^2 + 2.59\bar{p} - 0.129 \qquad (2.6.3)$$

Although this function implies that F becomes negative when $\bar{p} < 0.056$ or > 0.864 in reality most populations do not go outside this boundary. Eq 2.6.3 will be used below in order to estimate the effective sample size with which to fit the cline.

Table 2.6.3 Distribution of genetic parameters across the Pešćenica transect. Sites are included where there are five or more individuals. Ne is effective sample size (when F_{st} =0.025; see Chapter 3.4 for explanation). X/W is the distance from the centre of the cline standardised by the width; the sites are arranged in increasing order of this measure. p(obs) is the observed mean *variegata* gene frequency and p(exp) is the expected value from the model of the 2-D cline when habitat is not taken into account. pexp α is the expected gene frequency allowing for a deviation according to its habitat. A * indicates that the observed value deviates significantly from the model ($\Delta L>2$). F_{is} and R are the average deficit of heterozygotes and disequilibrium respectively. A * here indicates that the value is significantly different to 0 while the estimate in brackets is the value expected from the general expression for these two parameters (see text). The hybrid index is the number of individuals at each site with 0,1....8 *variegata* alleles. The majority of individuals were scored for 4 loci; where otherwise (3 or 5) the number has been scaled to 8.

									HYBRID INDEX								
site	Ne	hab	X/W	p(obs)	p(exp)	pexp α	Fis(exp)	R(exp)	0	1	2	3	4	5	6	7	8
2166	90	2	-2.72	0.069	0.08	0.11	0.161 (0.026)	0.062 (0.056)	25	12	5						
1	86	1	-2.71	0.080	0.08	0.06	0.000 (0.044)	-0.050 (0.078)	27	15	8						
1050	107	1	-2.51	0.039	0.09	0.06*	0.000 (-0.036)	0.219 (-0.020)	14	1	3						
1052	104	1	-1.83	0.063	0.09	0.07	0.294 (0.014)*	-0.050 (0.041)	22	12	4						
3	71	1	-1.74	0.059	0.09	0.07	0.000 (0.006)	0.400 (0.032)	13	4	1	1					
1053	83	1	-1.64	0.037	0.09	0.07*	0.158 (-0.041)	0.145 (-0.025)	13	3	1						
2120	22	2	-1.47	0.114	0.10	0.12	0.216 (0.101)	0.227 (0.150)	3	2		1					
2119	34	2	-1.41	0.071	0.10	0.12	0.101 (0.030)	0.093 (0.061)	4	2	1						
4	71	1	-1.35	0.068	0.10	0.07	0.000 (0.024)	-0.050 (0.053)	11	10	1						
1039	105	1	-1.27	0.062	0.10	0.07	0.196 (0.011)	0.000 (0.038)	38	17	3	2					
2	97	1	-1.26	0.067	0.10	0.07	0.000 (0.021)	-0.050 (0.050)	29	15	5						
2121	53	2	-1.01	0.020	0.10	0.13*	0.000 (-0.079)	-1.000 (-0.071)	7	1							
1040	89	1	-1.01	0.068	0.10	0.07	0.000 (0.023)	0.059 (0.052)	26	9	2	3					
2116	56	1	-1.01	0.081	0.10	0.07	0.000 (0.047)	-0.050 (0.082)	9	7	2						
1033	59	2	-0.97	0.069	0.10	0.13	0.130 (0.025)	0.206 (0.055)	10	3	1	1					
2135	35	2	-0.96	0.095	0.10	0.13	0.000 (0.071)	0.160 (0.112)	6	4		1					
5	57	1	-0.78	0.121	0.10	0.07	0.000 (0.112)	0.034 (0.164)	11	10	6	2					
1043	53	1	-0.71	0.126	0.11	0.08	0.000 (0.119)	-0.050 (0.174)	10	9	6	1	1				
10	16	1	-0.69	0.286	0.11	0.08*	0.000 (0.246)	0.185 (0.362)	1	1	2	2		1			
1035	75	1	-0.64	0.080	0.11	0.08	0.092 (0.046)	0.000 (0.080)	17	14	1	2					
1042	47	1	-0.63	0.056	0.11	0.08	0.000 (-0.001)	0.159 (0.023)	7	2		1					

Table 2.6.3 continued

									HYBRID INDEX								
site	Ne	hab	X/W	p(obs)	p(exp)	pexp α	Fis-exp	R(exp)	0	1	2	3	4	5	6	7	8
1014	76	1	-0.60	0.023	0.11	0.08*	0.000 (-0.073)	0.000 (-0.064)	9	2							
2143	55	2	-0.58	0.134	0.11	0.14	0.078 (0.130)	0.397 (0.188)*	19	12		1	2	1		1	
2115	29	2	-0.54	0.093	0.11	0.14	0.000 (0.068)	-0.050 (0.108)	3	3	1						
2152	64	1	-0.54	0.094	0.11	0.08	0.000 (0.071)	0.252 (0.111)	14	11		2	1				
1044	31	2	-0.52	0.236	0.12	0.15*	0.136 (0.226)	0.051 (0.324)*	5	3	5	3	3				
2145	16	2	-0.51	0.208	0.13	0.16	0.277 (0.207)	0.523 (0.295)*	3	1		1			1		
1019	13	2	-0.49	0.227	0.13	0.17	0.545 (0.220)*	0.400 (0.315)*	3	1						1	
1104	28	1	-0.48	0.202	0.14	0.10	0.491 (0.202)*	0.620 (0.287)*	7	2	2					1	1
2117	48	2	-0.48	0.075	0.14	0.17	0.000 (0.036)	-0.013 (0.068)	7	5		1					
1066	20	2	-0.47	0.219	0.14	0.18	0.127 (0.215)	0.192 (0.306)	2	1	3	1	1				
2159	37	1	-0.46	0.134	0.14	0.11	0.091 (0.130)	0.292 (0.188)*	6	5	1		2				
1013	24	0	-0.45	0.181	0.15		0.267 (0.183)	0.628 (0.260)	5	2		1				1	
1045	39	1	-0.44	0.087	0.16	0.12	0.276 (0.059)	-0.050 (0.097)	4	5	1						
2151	34	2	-0.40	0.097	0.17	0.22	0.203 (0.075)	0.276 (0.117)	5	1	3						
1103	24	2	-0.39	0.227	0.18	0.23	0.000 (0.220)	0.529 (0.315)	6		1	1	1	1	1		
1069	17	2	-0.36	0.150	0.20	0.25	0.000 (0.150)	-0.400 (0.215)	1	2	2						
1064	40	2	-0.36	0.252	0.20	0.25	0.250 (0.234)*	0.320 (0.338)*	9	7	7	5	1		1	3	
1063	39	1	-0.33	0.301	0.22	0.16	0.180 (0.250)*	0.261 (0.370)*	9	6	2	8	6	5	1		
1055	50	1	-0.29	0.139	0.24	0.19*	0.377 (0.136)*	0.430 (0.196)*	15	6	2	1		1		1	1
1110	28	2	-0.28	0.577	0.25	0.31*	0.282 (0.152)	0.445 (0.314)	1	1	1	1	1	3	1	3	1
1056	32	2	-0.23	0.218	0.29	0.36	0.392 (0.214)*	0.427 (0.306)*	9	3		2	2	5	1		1
1054	42	2	-0.21	0.594	0.30	0.37*	0.153 (0.140)	0.236 (0.300)*	2		2	2	8	3	3	5	3
6	61	1	-0.20	0.121	0.31	0.24*	0.022 (0.113)	0.079 (0.165)	13	14	4	4					
1011	30	2	-0.17	0.381	0.34	0.41	0.297 (0.248)*	0.564 (0.391)*	8	1		1	2	4	4		1
1113	17	2	-0.03	0.609	0.46	0.53	0.000 (0.131)	0.377 (0.289)			1	1		2		1	1
2147	20	2	0.05	0.449	0.54	0.61	0.510 (0.226)*	0.476 (0.382)*	3	1				3		3	
2146	27	2	0.09	0.792	0.57	0.64*	0.239 (0.018)	-0.050 (0.120)						1	3	1	1
13	17	0	0.14	0.604	0.61		0.041 (0.134)	0.324 (0.293)			1	1	1		1	2	
1049	18	2	0.14	0.640	0.62	0.69	0.053 (0.109)	0.373 (0.263)*			2			1	7	7	
1002	56	2	0.15	0.684	0.62	0.70	0.023 (0.079)	0.096 (0.223)*		1	1	1	5	9	7	10	2
7	21	0	0.18	0.528	0.65		0.079 (0.184)	0.281 (0.346)	1			2	2	1	2	1	
1003	66	2	0.25	0.740	0.70	0.77	0.000 (0.045)	0.062 (0.170)*	1			1	3	7	13	10	4

Table 2.6.3 continued

site	Ne	hab	X/W	p(obs)	p(exp)	pexp α	Fis-exp	R(exp)	HYBRID INDEX 0	1	2	3	4	5	6	7	8
8	39	2	0.35	0.800	0.73	0.79	0.215 (0.014)	0.180 (0.112)					2		3	2	3
9	24	2	0.56	0.800	0.76	0.81	0.000 (0.014)	0.481 (0.112)					1	1		1	2
14	27	2	0.57	0.585	0.76	0.81*	0.171 (0.147)	0.163 (0.308)		1		2	3	1	3	2	
1087	25	2	0.58	0.825	0.76	0.82	0.000 (0.005)	-0.100 (0.088)							3	1	1
2154	40	2	0.60	0.755	0.76	0.82	0.031 (0.036)	0.269 (0.156)*		1		1		2		7	3
2150	13	0	0.60	0.222	0.76		0.000 (0.217)	0.378 (0.310)	2		1	1	1				
1001	66	2	0.65	0.730	0.77	0.82	0.053 (0.051)	0.103 (0.180)*		1	1	2	3	10	8	10	7
1071	30	2	0.67	0.900	0.77	0.82	0.396 (-0.010)	0.142 (0.021)						1		1	3
1100	30	2	0.67	0.833	0.77	0.83	0.462 (0.002)	0.013 (0.081)						1	2	1	2
1070	39	2	0.77	0.820	0.78	0.83	0.000 (0.007)	0.303 (0.094)				1		1		5	2
12	16	2	0.84	0.650	0.79	0.84	0.000 (0.102)	0.110 (0.254)*				1		2	1	1	
2156	51	2	1.06	0.854	0.81	0.86	0.165 (-0.004)	-0.015 (0.061)						1	4	3	4
2158	27	2	1.15	0.850	0.82	0.86	0.045 (-0.003)	0.376 (0.065)					1			2	2
18	81	2	1.15	0.885	0.82	0.86*	0.000 (-0.009)	0.226 (0.034)					1	2	1	11	10
1074	94	2	1.18	0.850	0.82	0.87	0.000 (-0.003)	0.030 (0.065)				1		3	14	17	15
15	125	2	1.20	0.932	0.82	0.87*	0.025 (-0.009)	0.000 (-0.004)						1	7	18	38
16	50	0	1.22	0.781	0.82		0.000 (0.023)	0.267 (0.130)			1		1	3		9	2
1099	72	2	1.22	0.792	0.82	0.87	0.096 (0.018)	0.179 (0.120)		2				3	9	14	5
17	101	2	1.42	0.918	0.84	0.88*	0.000 (-0.010)	0.000 (0.007)							4	15	16
1097	31	2	1.50	0.854	0.85	0.89	0.000 (-0.004)	0.276 (0.061)						1	2		3
2126	36	2	1.52	0.828	0.85	0.89	0.000 (0.004)	-0.400 (0.086)							4	3	1
2140	24	2	1.53	0.773	0.85	0.89	0.000 (0.027)	0.226 (0.139)					1		2	3	
2134	45	2	1.53	0.886	0.85	0.89	0.000 (-0.009)	0.326 (0.033)						1		5	3
2200	54	2	1.53	0.936	0.85	0.89*	0.634 (-0.008)	-0.050 (-0.008)							2	1	7
2133	46	2	1.54	0.922	0.85	0.89	0.000 (-0.009)	0.019 (0.003)							1	3	4
20	71	2	1.59	0.955	0.85	0.89*	0.000 (-0.004)	0.134 (-0.022)							1	3	10
19	111	2	1.76	0.919	0.87	0.90*	0.000 (-0.010)	0.000 (0.006)							4	22	20
2163	39	2	2.03	0.893	0.88	0.91	0.000 (-0.010)	-0.200 (0.027)							1	4	2
2165	77	2	2.29	0.929	0.90	0.93	0.000 (-0.009)	-0.050 (-0.003)							1	8	9
1112	31	2	2.69	0.854	0.91	0.94	0.322 (-0.004)	0.476 (0.061)						2		1	3
1029	147	2	2.96	0.984	0.93	0.95	0.376 (0.006)	0.000 (-0.042)							1	3	33
1028	55	2	2.96	0.983	0.93	0.95*	0.000 (0.005)	-1.000 (-0.041)								1	5

Table 2.6.4 The maximum likelihood of $\bar{p}$, F_{IS} and R (the standardised disequilibrium) summed across populations of similar $\bar{p}$. N is the number of populations pooled together and $\bar{p}'$ is the mean frequency for that group. Limits are given in brackets and values which differ significantly from 0 are marked with an *. ΔL between loci or sites gives an estimate of the heterogeneity between loci or sites for that category. There are four degrees of freedom for ΔL between loci; for sites it is N-1. Significant differences are marked with an *

N	$\bar{p}'$	F_{IS} (limits)	ΔL between loci	ΔL between sites	R (limits)	ΔL between loci	ΔL between sites
23	0.068	0.040 (0.000, 0.097)	0.42	8.37	0.041 (0.000, 0.087)*	2.03	12.00
17	0.160	0.145 (0.073, 0.220)*	2.94	11.12	0.254 (0.231, 0.276)*	8.08*	43.40*
4	0.301	0.215 (0.103, 0.328)*	1.22	0.84	0.349 (0.283, 0.405)*	4.09	7.68*
2	0.487	0.286 (0.048, 0.505)	3.39	1.75	0.391 (0.232, 0.505)*	2.06	1.62
8	0.626	0.076 (0.000, 0.176)	0.84	2.31	0.237 (0.166, 0.306)*	5.09*	8.43*
13	0.771	0.016 (0.000, 0.101)	0.03	3.19	0.170 (0.110, 0.209)*	5.42*	12.19*
18	0.914	0.000 (0.000, 0.067)	1.14	5.20	0.000 (0.000, 0.068)*	1.75	14.44

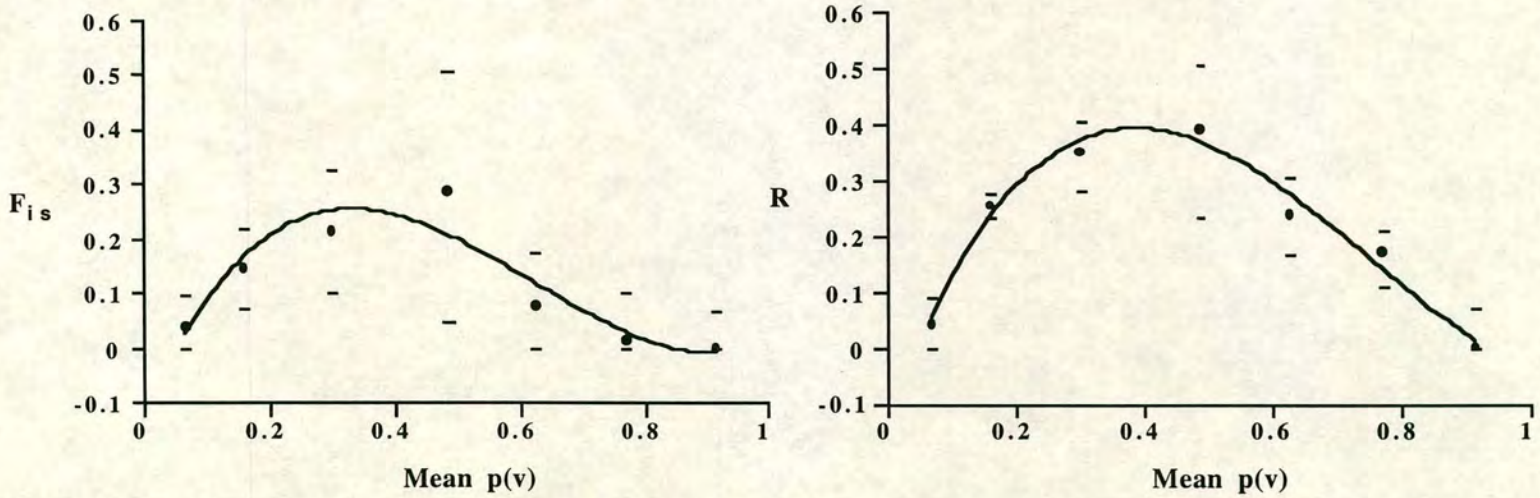


Fig. 2.6.2 General expressions for how F_{IS} and R change with gene frequency. Both increase towards the centre of the zone. Note the asymmetry; there is a sharper rise in F_{IS} and R on the *bombina* side of the zone (see Table above and sections 2.6 and 2.7).

2.7 Associations between loci

Method

Non-random associations between loci are referred to as linkage disequilibria. Linkage disequilibrium is defined directly from the gamete frequencies. For two alleles A and B at two loci;

$$D_{AB} = P_{AB} - P_A P_B \qquad (2.7.1)$$

where D is linkage disequilibrium, P_{AB} is the observed gamete frequency and $P_A P_B$ is the product of the gene frequencies. Although the name implies a physical link between the loci this need not be so. In the context of hybrid zones, linkage disequilibrium refers to the extent to which alleles from one population remain together despite recombination with another population (Chapter 1). Estimates of disequilibrium here are not measured directly from gametes but are inferred from the genotype of the individual. This means that estimating the contribution to disequilibrium of the double heterozygote for any pair of loci is difficult. An estimate that allows for this is therefore required.

Linkage disequilibrium plays a central role in the understanding and description of the genetic interactions between populations (Chapter 1). It may arise as a result of a number of forces such as selection, drift or migration. Although it is understood as the association between loci it can be estimated in various ways (Hedrick, 1987). Most of the discussion has centred on the ability to measure disequilibria between particular pairs of loci, (Golding, 1984; Hedrick, 1987; Hill, 1974; Hill, 1975; Hill & Robertson, 1968; Lewontin, 1988). This analysis involves measuring disequilibria among five loci i.e. ten pairwise values (Ak, Mdh, Ldh, Idh and Ck; Gpi is excluded as it is not diagnostic). The aim is to obtain an overall estimate of the strength of pairwise disequilibrium averaged across all the loci and see how this varies across the hybrid zone. This value will be used to estimate the rate of dispersal and strength of selection within the zone (Chapter 6). The overall strength of disequilibrium will also affect the effective sample size of populations with which to fit the cline (section 2.8).

Obtaining an overall estimate of disequilibrium is problematic. The strength of

disequilibrium depends on the allele frequencies from which it is derived (Lewontin, 1988; Hedrick, 1987). As long as this dependence on allele frequencies remain it is difficult; a) to compare disequilibria among different loci within a population, b) to compare disequilibria between the same loci within a population and c) to measure the cumulative value of disequilibria across a number of loci. A number of different solutions to allow for variation in allele frequencies have been proposed though none of them is satisfactory (Lewontin, 1988). The technique followed in this analysis is to standardise D by the allele frequencies as follows:

$$R = \frac{D}{\sqrt{pquv}} \qquad (2.7.2)$$

This is equivalent to a correlation if the loci are assigned states 0 or 1:-

$$R = \frac{cov(A,B)}{\sqrt{\sigma_A^2 \sigma_B^2}} \qquad (2.7.3)$$

Where the variance is $\sigma_A^2 = p_A q_A$ and $\sigma_B^2 = p_B q_B$. Even this measure however still depends on allele frequency as it will only be able to range fully from -1 to +1 when p = u = 0.5.

This is a measure of the average **pairwise** disequilibrium and ignores all other higher order associations. Strong pairwise disequilibria necessitate higher order associations so that negative gamete frequencies are not generated. These will not be estimated for a number of reasons. Primarily, it is not practical to evaluate all possible disequilibrium values as this would take an inordinate amount of computer time. Also, the sample sizes required to detect these disequilibria would have to be higher than this data s*et al.*lows. More importantly the information obtained by estimating the pairwise disequilibria values alone is sufficient to clarify the observed patterns that are seen in a tension zone and obtain those parameters we are interested in, essentially those of selection and dispersal.

There are two alternative methods for estimating the total disequilibrium value. It can either be measured as R using maximum likelihood (Hill, 1974) or, more simply from the variance in the "hybrid index" (Barton and Gale, 1993). A comparison of these two methods shows that the maximum likelihood method is more accurate and it is therefore used in this analysis (MacCallum, in prep).

In a similar manner to the calculation for F_{is} a support curve can be generated for

possible values of R. The rationale is as follows; assuming that gametes combine at random the expected genotype frequencies can be calculated for any value of R. That value of R which gives the data with the greatest probability will be that which maximises the log likelihood. This is chosen as the one that best explains the data.

It is known from the previous section that some sites show a significant heterozygote deficit. This has to be taken into account when estimating D as a consistent lack of heterozygotes across all loci would inflate any associations between loci. In order to overcome this it is assumed that for all populations a proportion (1-F) of gametes combine at random while the fraction F remain homozygous at both loci. F_{is} is estimated as the average for each pair of loci. Therefore for two loci (A, B) the expected frequencies of each genotype given a certain value of D and F are as follows:-

Genotype	Frequency
AABB	$(p_Ap_B+D)^2(1-F)+F(p_Ap_B+D)$
AaBB	$2(p_Ap_B+D)(q_Ap_B-D)$
aaBB	$(q_Ap_B-D)^2(1-F)+F(q_Ap_B-D)$
AABb	$2(p_Ap_B+D)(p_Aq_B-D)$
AaBb	$2(p_Ap_B+D)(q_Aq_B+D)+2(p_Aq_B-D)(q_Ap_B-D)$
aaBb	$2(q_Ap_B-D)(q_Aq_B+D)$
AAbb	$(p_Aq_B-D)^2(1-F)+F(p_Aq_B-D)$
Aabb	$2(p_Aq_B-D)(q_Aq_B+D)$
aabb	$(q_Aq_B+D)^2(1-F)+F(q_Aq_B+D)$

If N is the observed number for each genotype then the log likelihood value is:-

$$
\begin{aligned}
\text{Log}(L) = \quad & N_{(ABAB)}\log[(p_Ap_B+D)^2(1-F)+F(p_Ap_B+D)]+ \\
& N_{(ABaB)}\log[2(p_Ap_B+D)(q_Ap_B-D)] + \\
& N_{(aBaB)}\log[(q_Ap_B+D)^2(1-F)+F(q_Ap_B+D)] + \\
& N_{(ABAb)}\log[2(p_Ap_B+D)(p_Aq_B-D)] + \\
& N_{(ABab/aBAb)}\log[2(p_Ap_B+D)(q_Aq_B+D)+2(p_Aq_B-D)(q_Ap_B-D)] + \\
& N_{(aBab)}\log[2(q_Ap_B-D)(q_Aq_B+D)] + \\
& N_{(AbAb)}\log[(p_Aq_B-D)^2(1-F)+F(p_Aq_B-D)] + \\
& N_{(Abab)}\log[2(p_Aq_B-D)(q_Aq_B+D)] + \\
& N_{(abab)}\log[(q_Aq_B+D)^2(1-F)+F(q_Aq_B+D)] \qquad (2.7.4)
\end{aligned}
$$

D is standardised as R and different values of R are iterated to find the maximum likelihood value (F - values used are those observed for each population; section 2.6). For any population either D or R could be used as an estimate of association between a particular pair of loci. To obtain an overall estimate or to compare estimates between different pairs of loci R is used. The likelihoods of different standardised pairwise disequilibria (R) are summed together to give an overall estimate of disequilibrium across all pairs of loci. As each pairwise value is not independent (each locus forms a part of more than one pair) this estimate of R tends to be noisier than otherwise. Cross locus disequilibrium is dealt with by assigning gametes from the double heterozygote class in proportion to the frequency of the homozygote classes (i.e in proportion to their likelihood). An alternative method would be to divide the number of double heterozygotes equally into the four homozygote classes available (aabb, aaBB,AAbb and AABB). The likelihood approach is more accurate as it allows for differences in the frequency of cis and trans heterozygotes. This will be expected in a hybrid zone where each side of the zone will have a higher proportion of one genotype or the other.

Results

R was estimated for each pair of loci across all sites (Table 2.7.1). R differs significantly between pairs of loci (L_9 = 10.03; $p<0.05$). A significant difference between loci is not expected for neutral markers. This significance may be spurious as the pairwise comparisons are not independent of each other. The overall maximum likelihood estimate summed across all loci for all sites is R = 0.175. This estimate however is not very informative as it does not describe how disequilibrium varies

Table 2.7.1 Pairwise linkage disequilibria across all sites. Each value in the centre of the matrix is the maximum likelihood estimate of R, the standardised disequilibrium. The total maximum likelihood estimate of R summed across all loci is given at the bottom of the right hand column. The average for each locus is given in the rest of this column. The data come from 85 populations each with 5 or more individuals. The sample size for each locus varies as not all individuals were scored for every locus (see Appendix 2.2). Limits are given in brackets.

	Ak	Mdh	Ldh	Idh	Average
Ak					0.129
Mdh	0.176(0.134-0.207)				0.182
Ldh	0.157(0.099-0.212)	0.193(0.151-0.226)			0.165
Idh	0.180(0.110-0.252)	0.219(0.141-0.262)	0.235(0.161-0.302)		0.259
Ck	0.000(0.000-0.124)	0.141(0.014-0.266)	0.071(0.000-0.171)	0.403(0.000-0.0468)	0.154
					0.175

Heterogeneity between loci $\Delta L_9=10.03$; $p<0.005$
Heterogeneity between sites $\Delta L_{84}=152.37$; $p<0.001$

across the zone. R summed across all loci varies significantly between sites (ΔL_{84} = 152.37; $p<0.001$). It is difficult to see a consistent pattern as there are only a few individual sites where disequilibrium values differ significantly from zero (Table 2.6.3); individual sample sizes are small.

How does disequilibrium vary with gene frequency? As linkage disequilibrium is generated through a balance of dispersal of the parental genotypes into the zone and selection against recombinants within the zone it is expected that R will reach a peak in the centre where there is an intermediate gene frequency. Associations between loci break down under recombination (by 1/2 each generation); as combinations of alleles travel from one side of the zone to the other they are exposed to increased recombination. Introgressed alleles either side of the zone will therefore show less association between loci than those in the centre (details in Chapter 1). Disequilibrium will vary between sites depending on local selection pressures and immigration rates. Deviations from the local average may reveal how selection and dispersal interact at a local level. To estimate the expected value of R, populations within a certain range of $\bar{p}$ can be pooled together as for F_{is}. This will increase the effective sample size for each estimate thus providing a more accurate representation of the value of disequilibrium for any particular gene frequency (Table 2.6.4).

R differs significantly from 0 for each group. It increases for those populations with an intermediate gene frequency i.e. towards the centre of the zone. In populations whose gene frequencies are near but not at the extremes (where $\bar{p}'$ = 0.160, 0.626 and 0.771) there are significant differences between R for different pairs of loci. Towards the edges of the zone where there are fewer introgressed alleles not only will the total disequilibrium be less (having undergone more recombination) but so will selection against those recombinants. Associations between marker loci are held together by disequilibrium with other selected loci. As selection decreases higher order associations break down as well. If different assocations decay at different rates then differences in R between loci may be generated.

There are also significant differences in R between sites for all except the two purest groups of populations and the central populations ($\bar{p}'$ = 0.487). As disequilibrium is generated and maintained by dispersal and selection then the reason for differences in R between sites may be due to local differences in these forces. Dispersal and selection may change in relation to the underlying habitat of a particular population. This argument will be explored later.

A cubic polynomial curve best describes these data (Fig. 2.6.2); the variation accounted for is significantly greater than that explained by a quadratic fit ($F_{1,4} = 8.25$; $p<0.05$) but is not significantly improved by a quartic fit ($F_{1,3} = 0.43$). This provides a general expression for how disequilibrium is expected to vary with gene frequency:-

$$R = 2.29\bar{p}^3 - 5.25\bar{p}^2 + 3.03\bar{p} - 0.13 \qquad (2.7.5)$$

The results show a similar pattern to that of F_{is}.

1. Disequilibrium is strong at its peak; R = 0.388 when $\bar{p} = 0.38$.
2. The graph is asymmetric. There is a steeper rise in R for *bombina* -like populations than there is for *variegata* -like ones.

The fact that R is strong at its peak suggests either that dispersal is large and/or selection against recombinants is strong. The relative strengths of dispersal and selection will be estimated later when the width and shape of the cline is known (Chapter 6).

Synopsis of results

All the analysis so far has been done in relation to gene frequency either for each locus separately or in relation to $\bar{p}$. Three distinct results emerge from this.

1.The shift in gene frequency occurs in parallel for each locus.

2.There is strong linkage disequilibrium between these loci and a large and similar heterozygote deficit within loci. Both these parameters peak for populations of intermediate gene frequencies.

3.The polynomials describing R and Fis are asymmetric in shape; both forces are stronger for *bombina* like populations.

2.8 Fitting a cline in two dimensions

A cline fitted in one dimension across a hybrid zone makes the assumption that contours of gene frequency are linear. This is generally not the case as clines are not always straight and their form is not constant. This section will demonstrate how a cline can be fitted in two dimensions by maximum likelihood. Once the most appropriate position of the cline centre is found the data can be reduced to one dimension allowing for a detailed analysis of cline shape.

Finding the most likely centre of the cline is not trivial, especially in two dimensions. There are three aspects to this problem, all of which will affect the likelihood of the outcome; the source and amount of error between the observed and expected gene frequency for any population, the nature of the model generating the expected frequency and finally how the maximum likelihood is estimated when many variables are involved. Each of these three issues will be described and discussed in turn. The results will be presented at this stage. Finally the data will be reduced to one dimension and the adequacy of the fit assessed.

2.8.1 Estimating the sampling error

No sound estimate of how well a model describes the data can be made without specifying the error. The cline is fitted using the mean *variegata* frequency for each population ($\bar{p}$). There are at least two known sources of variation within this data set; random fluctuations in allele frequency between sites due to drift and deviations between sites generated by actual sampling.

The effect of fluctuations in actual frequencies becomes increasingly important as the sample sizes of populations increase. If underlying fluctuations are not accounted for, the positioning of the cline will tend to be dominated by the large samples. Any deviation in these sites will a have an appreciable effect and may skew the position of the cline unrealistically. Sites with small sample sizes contain information about the course and shape of the cline even where there is only one individual. It is therefore necessary to account for any increase in the variance of $\bar{p}$. The variance will be affected by discordance between loci, disequilibrium between loci and any heterozygote deficit within loci.

It is assumed here that any discordance between loci is caused by drift. This will also be reflected by a discordance of $\bar{p}$ from its expected value on the cline. This can be described and estimated as F_{st}' where F_{st}' is estimated from the variation across all the diagnostic loci around their mean (section 2.5). The sample sizes of the populations can be reduced in proportion to the effect any deviation may have in the following way.

$$\frac{1}{N_e} = \frac{1}{N} + \frac{Fst'}{k} \qquad (2.8.1)$$

(Szymura and Barton, 1991; note incorrect bracketing in original). Where N is the total number of genes, k is the number of loci and N_e is the effective sample size (Table 2.8.1). This therefore has a greater effect on larger samples. $F_{st}' = 0.0068$ (section 2.5) so if N = 10, $N_e = 8.5$ (an 18% reduction approximately) but if N = 200 then $N_e = 149.3$ (a 25% reduction). This does not take into account any concordant changes between loci which may arise through local selection or habitat differences and will therefore be an underestimate of the true error (Chapter 3).

Disequilibrium between loci causes an increase in the sampling variance of $\bar{p}$. If each locus were independent then the variance of the mean would be $\Sigma_i p_i q_i/(2k^2N)$; note that N here is the number of diploid individuals (unlike eq.2.8.1, 2.8.2 and 2.8.3). As the loci do not vary independently the variance at this mean will be increased by $\Sigma_i\Sigma_j D_{ij}/(2k^2N)$ where D_{ii} is defined as $p_i q_i$, (Sites *et al.*, 1994). For populations of intermediate frequency i.e. the central populations the effect will be large.

F_{is} also has an effect on the sampling variance of $\bar{p}$ as a deficit of heterozygotes will increase any estimate of disequilibrium (Section 2.5). Overall, assuming F_{is} is constant across loci and R is equal across all pairs of loci

$$\mathrm{var}(\bar{p}) \equiv \frac{\bar{p}\bar{q}}{2N_e k^2} = \sum_i \sum_j \left[\frac{(1+F_{is})D_{ij}}{Nk} \right] \qquad (2.8.2)$$

This becomes:-

$$N_e = \frac{N}{(1+F_{is})(1+R(k-1))} \qquad (2.8.3)$$

(N.Barton, pers comm), where N is the total number of genes and D is given as R, the standardised disequilibria averaged across all pairs of loci (Table 2.7.1). Variation in allele frequency across loci is neglected.

Table 2.8.1 The effective sample size for each population used to fit the cline in two dimensions. X and Y are the co-ordinates of each site measured from the global origin (see section 2.8.2). $\bar{p}$ is the mean *variegata* frequency averaged across the diagnostic loci. The total number of genes and diagnostic loci scored are given for each site (lumped between years). Ne is the effective sample size when random fluctuations are taken into account. Ne* is the subsequent sample size once disequilibria and deviations from Hardy-Weinberg are allowed for. See section 2.8.1 for details.

Site	X	Y	$\bar{p}$	Number of: genes	loci	Ne	Ne*
1	10.69	12.24	0.079	392	4	235.24	183
2	1.04	9.79	0.067	390	4	234.52	200
3	5.54	6.39	0.059	152	4	120.79	110
4	6.39	3.54	0.068	176	4	135.47	114
5	4.84	1.74	0.121	232	4	166.38	100
6	3.39	-0.00	0.121	280	4	189.70	114
7	1.94	-0.31	0.528	72	4	64.15	27
8	1.54	-1.11	0.800	80	4	70.42	52
9	1.39	2.09	0.800	40	4	37.45	28
10	1.54	2.64	0.286	56	4	51.13	20
11	-1.81	1.24	0.042	24	4	23.06	25
12	-1.21	2.39	0.650	40	4	37.45	19
13	-3.36	1.34	0.604	48	4	44.38	21
14	0.34	-1.96	0.585	94	4	81.05	37
15	-4.06	-2.61	0.932	512	4	273.74	280
16	-2.56	-3.51	0.781	128	4	105.15	74
17	-3.21	-4.46	0.918	280	4	189.72	188
18	-1.76	-5.36	0.885	200	4	149.25	137
19	-2.71	-8.11	0.919	368	4	226.38	225
20	-5.91	-5.91	0.955	112	4	94.09	101
1001	-0.47	-1.73	0.730	318	4	206.41	128
1002	-1.19	-0.95	0.684	272	4	186.00	103
1003	-0.93	-1.05	0.740	296	4	196.91	125
1004	1.94	-0.31	0.455	22	4	21.21	8
1005	3.15	-0.11	0.125	8	4	7.89	5
1010	2.59	0.60	0.250	8	4	7.89	3
1011	2.74	0.83	0.381	168	4	130.68	48
1013	3.71	1.19	0.181	72	4	64.15	30
1014	4.70	0.87	0.023	88	4	76.55	102
1016	2.45	1.06	0.042	24	4	23.06	25
1018	3.32	1.39	0.500	8	4	7.90	3
1019	4.08	0.93	0.227	44	5	41.52	15
1025	-8.00	-16.00	1.000	8	4	7.90	9
1028	-8.00	-16.00	0.983	60	5	55.47	66
1029	-8.00	-16.00	0.984	322	5	223.94	267
1032	6.13	4.09	0.000	10	5	9.87	24
1033	5.71	-5.03	0.069	116	5	100.19	80
1035	3.95	-5.12	0.080	250	4	175.44	135
1036	4.08	-5.09	0.062	16	4	15.58	14
1037	1.32	0.80	0.050	20	5	19.47	19
1038	1.20	0.61	0.367	30	5	28.82	9
1039	8.32	0.69	0.062	454	4	256.24	228
1040	6.93	0.91	0.068	310	4	203.01	172
1041	0.37	-0.60	0.600	10	5	9.87	4
1042	3.97	-5.03	0.056	72	4	64.19	60
1043	4.60	-4.82	0.126	206	4	152.57	90
1044	4.07	-4.28	0.236	144	4	115.68	48

Table 2.8.1 continued (2)

Site	X	Y	$\bar{p}$	Number of: genes	loci	Ne	Ne*
1045	3.73	-4.17	0.087	80	4	70.42	52
1046	3.78	-4.19	0.458	24	4	23.06	9
1047	3.31	-4.12	0.437	16	4	15.57	6
1049	2.32	-0.86	0.640	50	5	46.82	21
1050	10.57	8.05	0.039	180	5	144.60	163
1051	10.94	8.19	0.150	20	5	19.47	9
1052	10.98	0.11	0.063	318	5	221.99	188
1053	-1.00	15.00	0.037	136	4	110.46	125
1054	2.63	-4.46	0.594	212	4	155.838	72
1055	2.84	-4.62	0.139	202	4	150.37	83
1056	2.63	-4.59	0.218	142	4	114.39	49
1057	4.34	0.74	0.000	24	3	22.77	35
1058	2.53	-0.89	1.000	10	5	9.87	12
1059	0.72	-1.61	0.767	30	5	28.82	18
1060	2.06	-5.00	0.750	8	4	7.89	5
1061	4.34	-4.99	0.187	16	4	15.58	7
1063	2.93	-5.58	0.301	296	4	196.91	75
1064	3.05	-4.69	0.252	262	4	181.26	73
1066	6.27	-7.37	0.219	64	4	57.72	25
1067	-0.47	-1.73	0.583	24	4	23.06	10
1068	-0.47	-1.73	0.625	8	4	7.89	4
1069	3.06	-5.71	0.150	40	4	37.45	20
1070	1.64	-7.56	0.819	72	4	64.15	50
1071	1.63	-7.27	0.900	40	4	37.45	36
1072	0.22	-5.80	0.875	24	4	23.06	21
1073	1.61	-7.01	0.875	8	4	7.89	7
1074	-0.28	-7.20	0.850	380	4	230.86	194
1075	-0.09	-4.03	0.875	16	4	15.58	14
1076	-0.60	-5.20	0.938	16	4	15.58	16
1077	-0.02	-5.43	0.889	18	3	17.29	16
1078	1.61	-7.21	0.781	32	4	30.35	21
1079	0.09	-5.80	0.833	12	3	11.68	10
1080	0.44	-5.90	0.750	8	4	7.89	5
1081	6.33	-7.48	0.125	32	4	30.35	18
1082	5.49	-6.74	0.250	32	4	30.35	12
1083	2.52	-6.69	0.625	8	4	7.89	4
1084	2.44	-6.60	0.250	24	4	23.06	9
1085	2.41	-6.52	0.417	24	4	23.06	9
1086	2.10	-6.62	0.750	8	4	7.89	5
1087	1.61	-6.99	0.825	40	4	37.45	29
1089	-0.75	-5.35	1.000	8	4	7.89	9
1091	-1.04	-5.68	0.938	32	4	30.35	31
1092	-1.72	-5.93	0.875	8	4	7.89	7
1097	-2.87	-6.14	0.854	48	4	44.38	38
1098	-2.07	-4.90	1.000	8	4	7.89	9
1099	-2.34	-4.68	0.792	264	4	182.22	132
1100	-2.92	-0.89	0.833	48	4	44.38	36
1103	2.77	1.70	0.227	88	4	76.55	32
1104	3.17	1.86	0.202	104	4	88.38	39
1105	2.97	1.85	0.167	24	4	23.06	11
1109	3.26	-4.78	0.000	8	4	7.89	15
1110	3.10	1.01	0.577	104	4	88.38	40
1111	-8.00	-15.00	1.000	16	4	15.58	18
1112	-8.00	-14.00	0.854	48	4	44.38	38
1113	1.82	0.41	0.609	46	4	42.67	20
2012	3.10	1.01	0.464	28	4	26.73	10

Table 2.8.1 continued (3)

Site	X	Y	$\bar{p}$	Number of: genes	loci	Ne	Ne*
2054	2.63	-4.46	0.125	24	4	23.06	14
2115	4.07	-4.41	0.093	54	4	49.46	35
2116	6.40	1.60	0.081	136	4	110.46	85
2117	3.23	1.82	0.074	94	4	81.05	65
2118	-0.60	-7.25	0.813	16	4	15.58	12
2119	7.86	-6.93	0.071	56	4	51.13	42
2120	7.96	-6.89	0.114	44	4	40.94	26
2121	7.02	-6.84	0.020	50	4	46.08	64
2122	-6.69	-5.88	0.938	32	4	30.35	31
2124	-6.51	-5.58	0.813	16	4	15.58	12
2126	-6.06	-5.43	0.828	64	4	57.72	46
2127	-5.80	-5.44	0.969	32	4	30.35	34
2132	-5.57	-5.27	0.938	16	4	15.58	16
2133	-5.22	-5.34	0.922	64	4	57.72	58
2134	-5.62	-5.39	0.886	70	4	62.56	57
2135	6.92	-6.84	0.095	74	4	65.73	46
2136	-5.93	-5.48	0.900	20	4	19.34	18
2138	-6.24	-5.58	0.875	8	4	7.89	7
2140	-5.81	-5.39	0.773	44	4	40.94	28
2141	-5.73	-5.37	0.846	26	4	24.90	21
2142	6.53	-1.94	0.167	24	4	23.06	11
2143	5.65	-0.71	0.134	246	4	173.46	98
2144	4.74	-1.83	1.000	4	1	3.89	4
2145	5.40	-1.30	0.208	48	4	44.38	20
2146	2.74	-1.26	0.792	48	4	44.38	32
2147	2.73	-1.78	0.449	78	4	68.87	26
2148	5.65	-1.61	0.125	32	4	30.35	18
2149	6.13	4.09	0.500	8	4	7.89	3
2150	0.28	-1.67	0.222	36	4	33.92	14
2151	2.92	-5.17	0.097	72	4	64.15	44
2152	5.44	-0.66	0.094	212	4	155.84	109
2153	6.07	-13.06	0.250	24	4	23.06	9
2154	2.07	-11.78	0.755	98	4	84.00	55
2155	2.08	-11.35	0.813	32	4	30.35	23
2156	3.41	-11.26	0.854	96	4	82.53	70
2157	3.94	-11.49	0.125	24	4	23.06	14
2158	3.80	-11.36	0.850	40	4	37.45	31
2159	4.30	-11.50	0.134	112	4	94.09	53
2163	-3.74	-8.60	0.893	56	4	51.13	48
2164	-5.76	-11.44	1.000	32	4	30.35	35
2165	-5.82	-11.29	0.930	142	4	114.39	116
2166	9.01	-12.43	0.069	332	4	212.22	177
2167	9.79	-12.42	0.031	32	4	30.35	36
2200	-5.80	-5.44	0.936	78	4	68.87	71

Both estimates can be combined by substituing N from eq 2.8.3 with N_e from equ 2.8.1. The effective sample size for each population was estimated using the value of R and F_{is} that would be expected given the mean *variegata* frequency for that population. (equation 2.6.3; equation 2.7.5).

The overall effective sample size incorporating drift, disequilibrium and any heterozygote deficit was estimated for each population (Ne*, Table 2.8.1). The mean frequency for each population and this effective sample size was used to fit the cline (Table 2.8.1). The data set used includes data on all individuals that had been scored for the allozymes. Sites containing only one individual can still provide information about the course of the cline. Therefore, unlike the analyses above where populations containing less than five individuals were excluded, all 147 populations are included in this analysis.

2.8.2 The model - a stepped cline

Within this hybrid zone there is evidence for both high concordance and strong disequilibrium between the neutral enzyme markers. This has important implications for choosing the appropriate model with which to fit the cline. The linkage disequilibria mean that selection at any one locus will cause changes at all the loci associated with it even though these may themselves be neutral. This results in an increase in the effective selection at any one locus. If selection acted independently at each locus then the shape of the cline would be expected to follow a smooth sigmoid curve (Endler 1977). However the increased effective selection implies that a stepped cline would be a more suitable model i.e. one where the rate of decay of introgressing alleles occurs over a much shallower gradient than would be expected from the step in the centre (Szymura and Barton, 1986). Evidence is provided later that the cline is indeed stepped. It is important that this is taken into account at this stage as otherwise undue weight would be given to those populations with a high frequency of introgressed alleles.

The model of a stepped cline consists of three parts; a central tanh curve ($P = 1+\tanh[2(x-y)/w]/2$]; this is equivalent to the exponential equation 1.1 given in Chapter 1) and two exponential tails where allele frequency declines at a rate $\exp(-4x\sqrt{\theta}/w)$. These form three straight lines when plotted on a logit scale [$z = \log_e(p/q)$].

The parameters defining cline shape in this model are as follows:-

1. The position of the centre of the cline (y).

2. The width of the cline (w). This is defined as 1/maximum gradient (see Chapter 1 for an explanation).

3. The rate of decay of introgressing alleles either side of the cline (θ_b, θ_v). This is equal to the square of the ratio between the scale over which the tail decays, and the width of the cline. It is proportional to the ratio between selection at an individual locus and the effective selection combined across all loci, i.e. the selection on the enzyme markers in the tail compared to that in the centre. If selection were the same then the ratio would = 1 and one would get a sigmoid curve.

4. The barrier to gene flow either side of the zone (B_b, B_v). This is defined as the ratio between the step in allele frequency in the centre of the zone and the gradient of the introgressing tail at the edge $B = \Delta p/(dp/dx)$ (Nagylaki, 1976, Chapter 1). This ratio has the dimensions of distance and is equivalent to the length of unimpeded habitat that would be the same obstacle to the flow of a neutral allele (Barton and Gale, 1993, Chapter 1 provides an explanation).

The expected frequency of any population will depend on its distance from the centre of the cline and the width of the cline at that point. The stepped cline is described by three curves; a central tanh curve and two exponential tails. Therefore for any population three values of the expected gene frequency can be estimated. The intermediate value of these three is generally the appropriate one to take.

Each site is described by a set of co-ordinates measured from a global origin (45° 38', 16° 10'E; Table 2.8.1). The centre of the cline is described by a linear chain of segments of equal length (Fig. 2.8.1). The course of the cline is defined by the angle between these segments. The distance of each site from the centre is the shortest distance to the cline. This may be to a corner or to a segment. The width of the cline is defined at each corner. If the width varies along the cline and therefore between segments then the width at any position along the cline is found by linear interpolation between the corners.

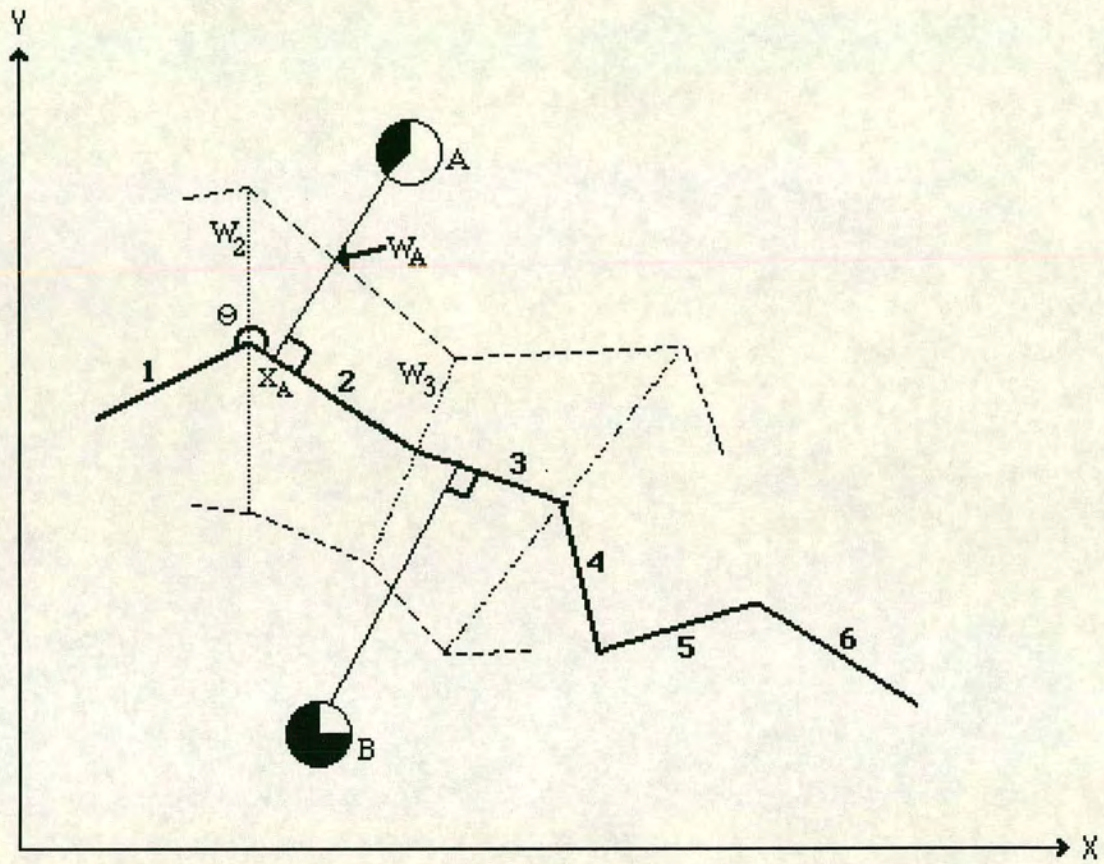


Fig. 2.8.1 Diagramatic representation of a cline in two dimensions described by six segments showing two populations, A and B. The proportion of *variegata* alleles is represented by the amount of black in each pie. θ describes the angle of direction between adjacent segments. The width of each segment is defined from the corners (at W_2 and W_3). The width at any point along a segment is found by linear interpolation between these. X_A is the position on the centre line (the centre of the cline) closest to A. The width of the cline at this point is W_A.

2.8.3 Statistical considerations: using the Metropolis algorithm

A large number of variables are used to define the centre and shape of the cline. The aim is to find the optimum values for all these parameters i.e. to find the cline that gives the data with the greatest probability. A different likelihood will be generated every time a variable is changed. The problem is to find the parameter set that maximises the likelihood.

Different solutions to multivariate optimisation have been put forward (Kirkpatrick *et al.* , 1983). There are two strategies;

1. The system can be divided into smaller more manageable problems. The optimum solution for each of these can then be pieced together to produce an optimum parameter set for the entire system in the hope that there is no cost due to interactions between the subdivided groups. This is known as the "divide and conquer" strategy.

2. Alternatively the system can be kept whole and parameters altered through iterative improvement. The system is rearranged from a known parameter set. A standard change is made to all the parameters of the system in turn until a configuration is produced with a lower cost. The process is repeated with this new improved configuration until no further improvement can be made.

The problem with both these approaches is that there may be only a very small chance of finding the global optimum. It is likely that many local optima can be found when there are many variables involved. The problem can be likened to an adaptive landscape, the aim being to find the highest peak in the presence of many inferior peaks. Kirkpatrick *et al.* devised a method based on the Metropolis algorithm to find the most likely solution. It incorporates both of the above strategies. This is the method that will be used to find the centre of the cline.

The Metropolis algorithm was intended to simulate the behaviour of atoms in equilibrium at a certain temperature (Metropolis *et al.* , 1953). It forms the basis of a useful analogy which can be applied to combinatorial optimisation. In the original algorithm a small random change is applied to a configuration of atoms. This results in a change in energy of the system. If this change is negative then the change is accepted. If the change is positive then the change is accepted with a probability equal

to $\exp(-\Delta E/k_B T)$ where k_B is Boltzmann's constant. If the atoms are replaced by a set of parameters and the change in energy by a change in the 'cost' between the original and changed configuration then this algorithm can be applied to any multivariate problem. A population of different configurations can be generated around the original by iterating the above procedure. Temperature is used as a control parameter such that the higher the temperature the more likely a change is accepted. The log likelihood ratio between two different cline fits is used as the "cost" in this analysis. The probability of accepting a change therefore becomes $\exp(+\Delta \log L/T) = L^{1/T}$ Therefore for any given temperature the Metropolis algorithm generates a random walk around the original parameter set with a density proportional to the likelihood raised to the power 1/T. As T approaches 0 then only uphill changes are made; if T = 1 then the density equals the likelihood.

This is the approach of strategy two; iterative improvement. However there is still the problem that the system will get stuck at a local optimum. Kirkpatrick *et al.* (1983) solved this by allowing the system to imitate annealing; a perfect crystal is formed from a melt by allowing it to cool slowly and spend a great deal of time at very low temperatures. If the cooling procedure is too fast then the crystal formed has many defects and only locally optimal structures. In a similar manner the Metropolis algorithm is allowed to undergo an annealing schedule. Large changes in the system are accepted at high temperatures. This means that different local optima become accessible; more of the landscape becomes visible. As the temperature is reduced the probability of accepting a change becomes smaller and the system settles at some peak. If the system is cooled infinitely slowly the global optimum will be reached. This is obviously not practical, so a compromise has to be made.

An advantage of the system is that gross features become resolved at higher temperatures where large changes are made and the more detailed aspects filled in later. This is essentially following the features of the first strategy i.e. that of 'Divide and conquer".

The cline was fitted using this system (using the program, 'analyse' written by N.H. Barton in Pascal to run on Macintosh computers). The temperature was gradually reduced by single degrees from 20, allowing 20 iterations at each step. The last step involves taking the best fit from all the iterations and constraining the system so that only uphill changes are allowed. When no further improvements are made (to three decimal places of log likelihood units) the process is stopped.

Each Metropolis run starts with a certain configuration of all the parameters. Replicate runs of the Metropolis algorithm can be made with the starting values for the model kept the same or different. Each run will produce a likelihood value which provides information about the fit of the model. This is the ratio between the likelihood of the observed model and and that expected from a perfect fit. The likelihood value will be approximately distributed as $\frac{1}{2}\chi^2_\nu$ where $\nu = N_{sites}$-df (number of sites, degrees of freedom). This will therefore give an estimate of that variation not accounted for by the model. The parameter set that is chosen will be the one to minimise the residual variation.

If the constrained parameters are changed then two different hypotheses of the same model can be compared. The difference in log likelihood between the two configurations will again be distributed as $\frac{1}{2}\chi^2_\nu$ but ν will now = df_1-df_2 where 1 and 2 are the two configurations involved. In this way different hypotheses concerning the two dimensional cline can be compared. Certain features of the model will be kept constant throughout all the Metropolis runs. These are the starting point of the cline and its length. As long as the starting point is set outside the sample area it should not affect the final position of the cline. The length of the cline should also be fixed and not allowed to extend beyond the data. This will not make much difference to the resulting likelihood value but there will be extraneous segments with no data to describe them giving a spurious increase in degrees of freedom.

Other features of the model can vary; the number of segments describing the length of the cline and whether the width varies or remains constant between segments: One linear segment through the data could describe the centre of the cline but this would not be very realistic; the cline is bound to curve. Increasing the number of segments will improve the fit but degrees of freedom will be sacrificed. A compromise therefore has to be reached. One requires the minimum number of parameters to realistically represent the cline. The number of segments should therefore be increased until there is no significant improvement. As the course of the cline is not expected to change direction over a series of sharp corners the cline can be approximated to a smooth curve once the optimum number of segments have been found. This 'smoothing' superimposes an increased number of segments onto the cline without altering the underlying cline shape and so does not alter the degrees of freedom of the original configuration. Essentially the sharpness of the angles between segments is decreased

by allowing the angle to change over more than one step. Smoothing is defined in multiples of the segment number, twice, three times and so on. If smoothing is twice the segment number then the angle changes over two steps, between segments, rather than one. Twice the segment length is generally appropriate (if segments are not too long) as this gives a reasonable approximation to a curve while not taking up too much computer time.

The width of the cline is allowed to vary but can either remain constant along the length of the cline or vary between segments. Allowing the width to vary between segments will again sacrifice degrees of freedom but may significantly improve the fit. Clines may vary in width along their length for a number of reasons (Chapter 1, Chapter 6.3). Clines maintained by a balance between dispersal and selection would not be expected to vary as much as those maintained by direct selection (Barton and Hewitt 1985). The simplest scenario therefore is to presume that the cline has a constant width and this is the initial presumption used here.

Results

As using the Metropolis algorithm takes a large amount of computer time an initial analysis was done which did not incorporate smoothing. This speeds up the time required for each replicate. Once a reasonable fit is found smoothing (at twice the segment length) can be incorporated. Fig. 2.8.2a shows a map of sites in relation to the global origin. The diagnostic gene frequency, $\overline{p}$, is represented as a pie for each site.

A constant length of 36km was set for the cline. The likelihood of clines composed of different numbers of segments were compared (Table 2.8.2a). For example the likelihood of a cline composed of 1 segment 36km long can be compared to 2 segments 18km long and so on up to 36 segments 1km long. In reality it was unrealistic to describe this cline with more than 12 segments as beyond this the segments were small enough to wrap around individual sites and effectively scrunch up. Even a cline described by 12 segments tended to curl up more often than not. Ultimately if there were enough segments the cline could wrap around all sites and give a perfect fit but it would be biologically unreasonable. Five runs were completed for a set number of segments. Although the starting position of the begining of the cline was kept constant (x,y are -10,-5 in relation to the global origin) the course of the cline beyond that point varied. The cline is best described with nine segments each

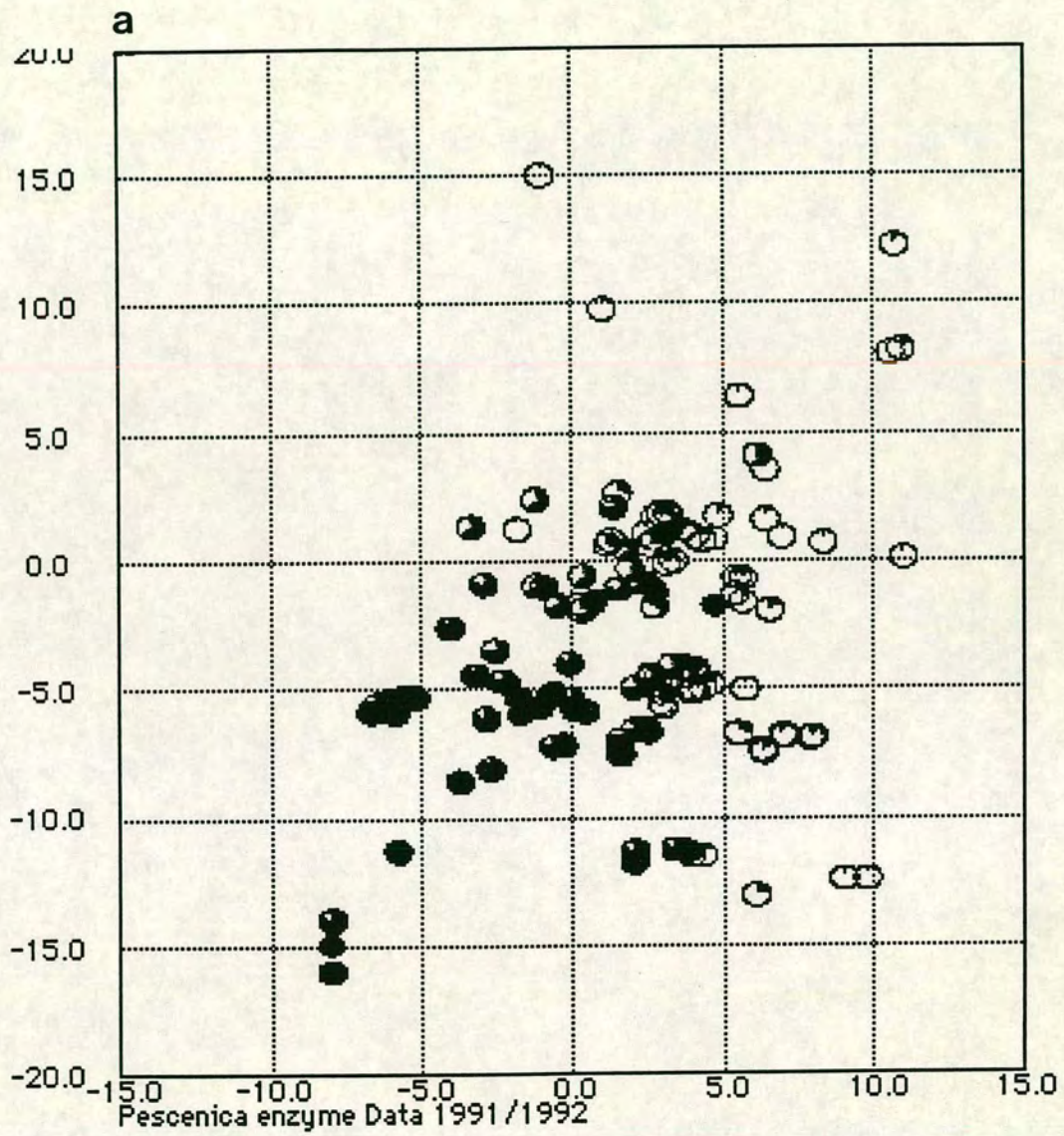


Fig 2.8.2 a The location and mean *variegata* gene frequency ($\bar{p}$), of populations sampled at Peščenica. Axes are measured relative to the global origin in km. **b**.The fitted cline in two dimensions. The width of the cline is constant. The centre of the cline is marked in red and the width in green (see text for details)

Table 2.8.2 a. Log likelihoods of clines composed of different numbers of segments. The total length of the cline remains the same at 36km. The runs are arranged in order of decreasing likelihood so that Run 1 is the best of the five replicates. The cline was fitted in two dimensions with 147 sites. The model was constrained to a constant width. The degrees of freedom = [number of sites-the four parameters describing the cline (Bb, Bv, θb, θv) - the number of widths (= segment no + 1) and the number of angles describing the segments -1]. The improvement made by increasing the number of segments can be assessed by comparing the difference in log likelihood between the best of the five runs given the difference in degrees of freedom. Although the best cline has 12 segments, 9 segments were chosen to describe the cline (log L= -266.19) as this gave a more reliable shape (see text, 2.8 for detail). **b**. The log likelihoods of the cline described by 9 segments with smoothing superimposed. Smoothing increases the likelihood of the cline from -266.93 to -251.96 with no loss in degrees of freedom.

a

No of segments	1	2	3	4	6	9	12
length (km)	36	18	12	9	6	4	3
Degrees of freedom	139	137	135	133	129	123	117
Run 1	-613.71	-565.07	-301.83	-295.82	-293.94	**-266.19**	-258.78
Run 2	-615.08	-618.62	-301.83	-301.95	-296.51	-267.99	-267.42
Run 3	-701.69	-621.77	-302.10	-313.08	-316.97	-269.56	-275.38
Run 4	-702.27	-628.04	-341.65	-354.55	-317.27	-283.70	-299.98
Run 5	-838.85	-644.44	-352.40	-367.87	-320.78	-286.48	-310.13

b. Cline: 9 segments, 4 km long - with smoothing

Run	log likelihood	Run	log likelihood
1	**-251.96**	6	-276.40
2	-255.92	7	-277.70
3	-265.65	8	-282.07
4	-272.60	9	-292.69
5	-272.96	10	-310.00

4km in length (L = 266.193). This is a significant improvement on six segments 6km in length ($\Delta L_6 = 27.75$; $p<0.001$) and although this is slightly improved by increasing the segment number to 12 ($\Delta L_6 = 7.41$), the addition of these extra segments tended to curl up the cline inappropriately.

If smoothing is incorporated into the cline described by nine segments such that each segment is represented by two (so smoothing = 18) then the likelihood of the cline improves by 14.23 units to 251.96 (Table 2.8.2b). There are no degrees of freedom sacrificed in this procedure as smoothing does not alter the underlying cline shape so the improvement is highly significant. Fig. 2.8.2b shows the the two dimensional cline superimposed onto the map of sites drawn from the global origin.

The cline can be viewed in one dimension by plotting the expected and the observed gene frequency as a function of the distance from the centre of the cline (Table 2.8.3, Fig. 2.8.3). This is plotted using a logit transformation ($z = \ln[p/q]$); if the cline followed a sigmoid curve the tails of the curve would then expand to form a straight line with the centre. The cline does not follow a straight line however, there is a sharp step in the centre flanked by two shallow tails. This indicates a strong barrier to gene flow (Chapters 1, 3). Details of cline shape will not be discussed in this chapter for reasons given below but the pattern of a stepped cline is expected given the strong disequilibrium in the centre of the zone (section 2.7). This increases the effective selection on any one locus which is in disequilibrium with other selected loci (Chapter 1). There is a large degree of scatter in the observed gene frequencies of populations around the expected cline shape.

The log likelihood of this cline is -251.96 with 123 degrees of freedom (df = N-[4 parameters estimating barrier strength and introgression rates either side of the zone + 10 describing the width for 9 segments + 9 describing the angles is 147-24]-1 = 123). If this model described most of the variation in gene frequency then the likelihood value should be less than 149/2 as the actual 5% level of $\chi^2_{123} = 149$. Therefore overall there is significant residual variation. This could be accounted for by significant deviations from the model at a few sites (Table 2.8.3). Out of 147 sites, 37 differed significantly from the model (= 25% of all sites-filled circles in Fig. 2.8.3). These subtract 200 units from the likelihood value. Of these 20 are significant at p = 0.01. Such a large amount of residual variation implies that the model used to describe the cline is inappropriate. The model used so far is that of a dispersal

Table 2.8.3 The observed and expected frequency of each site and its distance from the centre of the cline, X (Km). The logit transforrm (z = ln[p/q]) of the gene frequencies are given in brackets. Results are from fitting a cline in two dimensions assuming constant width along its length. The cline is approximated by nine segments 4km in length (with smoothing). The total likelihood of this cline is $L_{123} = 251.95$. * indicates that the observed frequency differs significantly from the expected at $p<0.05$; ** and $p<0.01$. N is the number of genes at each site. This is the effective sample size allowing for various sources of sampling error (see text for detail).

Site	N	p observed	p expected	X	Likelihood
1	183	0.08 (-2.46)	0.05 (-2.94)	-14.61	-1.54
2	200	0.07 (-2.64)	0.06 (-2.67)	-7.86	-0.00
3	110	0.06 (-2.77)	0.07 (-2.62)	-6.87	-0.07
4	114	0.07 (-2.61)	0.07 (-2.56)	-5.38	-0.02
5	100	0.12 (-1.99)	0.08 (-2.44)	-3.16	-1.00
6	114	0.12 (-1.98)	0.25 (-1.10)	-1.22	-5.84**
7	27	0.53 (0.11)	0.56 (0.22)	0.25	-0.04
8	52	0.80 (1.39)	0.69 (0.79)	0.88	-1.64
9	28	0.80 (1.39)	0.81 (1.45)	1.62	-0.01
10	20	0.29 (-0.92)	0.13 (-1.90)	-2.11	-1.68
11	25	0.04 (-3.13)	0.30 (-0.84)	-0.94	-5.49**
12	19	0.65 (0.62)	0.18 (-1.54)	-1.72	-10.43**
13	21	0.60 (0.42)	0.62 (0.49)	0.55	-0.01
14	37	0.59 (0.34)	0.81 (1.43)	1.59	-4.78**
15	280	0.93 (2.61)	0.88 (1.98)	2.88	-4.22**
16	74	0.78 (1.27)	0.88 (1.98)	2.87	-2.77*
17	188	0.92 (2.41)	0.89 (2.13)	3.97	-0.61
18	137	0.88 (2.04)	0.88 (2.03)	3.23	-0.00
19	225	0.92 (2.42)	0.91 (2.29)	5.17	-0.15
20	101	0.96 (3.06)	0.92 (2.47)	6.57	-0.90
1001	128	0.73 (0.99)	0.78 (1.27)	1.41	-0.91
1002	103	0.68 (0.77)	0.60 (0.40)	0.44	-1.60
1003	125	0.74 (1.05)	0.64 (0.56)	0.62	-3.03*
1004	8	0.45 (-0.18)	0.56 (0.22)	0.25	-0.16
1005	5	0.13 (-1.95)	0.30 (-0.86)	-0.96	-0.41
1010	3	0.25 (-1.10)	0.36 (-0.59)	-0.65	-0.08
1011	48	0.38 (-0.49)	0.31 (-0.79)	-0.87	-0.49
1013	30	0.18 (-1.51)	0.15 (-1.71)	-1.90	-0.09
1014	102	0.02 (-3.76)	0.08 (-2.43)	-2.74	-3.20*
1016	25	0.04 (-3.13)	0.34 (-0.66)	-0.73	-6.78**
1018	3	0.50 (0.00)	0.19 (-1.47)	-1.63	-0.75
1019	15	0.23 (-1.22)	0.13 (-1.95)	-2.17	-0.60
1025	9	1.00 (5.00)	0.96 (3.08)	11.38	-0.40
1028	66	0.98 (4.08)	0.96 (3.08)	11.38	-0.74
1029	267	0.98 (4.15)	0.96 (3.08)	11.38	-3.32*
1032	24	0.00 (-5.00)	0.07 (-2.56)	-5.50	-1.80
1033	80	0.07 (-2.60)	0.08 (-2.44)	-3.09	-0.07
1035	135	0.08 (-2.44)	0.16 (-1.70)	-1.88	-3.43*
1036	14	0.06 (-2.71)	0.14 (-1.81)	-2.01	-0.44
1037	19	0.05 (-2.94)	0.41 (-0.37)	-0.41	-6.56**
1038	9	0.37 (-0.55)	0.44 (-0.25)	-0.28	-0.09
1039	228	0.06 (-2.72)	0.07 (-2.59)	-6.11	-0.13
1040	172	0.07 (-2.62)	0.07 (-2.53)	-4.87	-0.05
1041	4	0.60 (0.41)	0.63 (0.55)	0.61	-0.00

Table 2.8.3 continued

Site	N	p observed	p expected	X	Likelihood
1042	60	0.06 (-2.83)	0.15 (-1.77)	-1.96	-2.49*
1043	90	0.13 (-1.93)	0.09 (-2.30)	-2.57	-0.62
1044	48	0.24 (-1.17)	0.09 (-2.27)	-2.52	-4.15**
1045	52	0.09 (-2.34)	0.12 (-2.00)	-2.23	-0.27
1046	9	0.46 (-0.17)	0.11 (-2.05)	-2.28	-3.34*
1047	6	0.44 (-0.25)	0.17 (-1.62)	-1.80	-1.23
1049	21	0.64 (0.58)	0.51 (0.06)	0.07	-0.67
1050	163	0.04 (-3.21)	0.06 (-2.82)	-11.45	-0.49
1051	9	0.15 (-1.73)	0.05 (-2.84)	-11.81	-0.55
1052	188	0.06 (-2.70)	0.06 (-2.70)	-8.56	0.00
1053	125	0.04 (-3.26)	0.05 (-2.84)	-11.94	-0.44
1054	72	0.59 (0.38)	0.26 (-1.03)	-1.14	-17.29**
1055	83	0.14 (-1.83)	0.23 (-1.20)	-1.34	-2.21*
1056	49	0.22 (-1.28)	0.26 (-1.03)	-1.15	-0.26
1057	35	0.00 (-5.00)	0.11 (-2.12)	-2.36	-3.97*
1058	12	1.00 (5.00)	0.47 (-0.11)	-0.12	-8.99**
1059	18	0.77 (1.19)	0.79 (1.30)	1.45	-0.02
1060	5	0.75 (1.10)	0.39 (-0.44)	-0.50	-1.33
1061	7	0.19 (-1.47)	0.12 (-2.04)	-2.27	-0.16
1063	75	0.30 (-0.84)	0.33 (-0.73)	-0.81	-0.11
1064	73	0.25 (-1.09)	0.21 (-1.31)	-1.46	-0.33
1066	25	0.22 (-1.27)	0.14 (-1.82)	-2.02	-0.57
1067	10	0.58 (0.34)	0.78 (1.27)	1.41	-0.97
1068	4	0.62 (0.51)	0.78 (1.27)	1.41	-0.25
1069	20	0.15 (-1.73)	0.32 (-0.74)	-0.83	-1.55
1070	50	0.82 (1.51)	0.79 (1.30)	1.45	-0.17
1071	36	0.90 (2.20)	0.76 (1.14)	1.26	-2.42*
1072	21	0.88 (1.95)	0.79 (1.30)	1.44	-0.57
1073	7	0.88 (1.95)	0.73 (1.01)	1.12	-0.42
1074	194	0.85 (1.73)	0.88 (1.96)	2.71	-0.60
1075	14	0.88 (1.95)	0.81 (1.43)	1.59	-0.23
1076	16	0.94 (2.71)	0.87 (1.86)	2.07	-0.44
1077	16	0.89 (2.08)	0.80 (1.37)	1.53	-0.47
1078	21	0.78 (1.27)	0.75 (1.12)	1.25	-0.04
1079	10	0.83 (1.61)	0.80 (1.40)	1.56	-0.03
1080	5	0.75 (1.10)	0.77 (1.20)	1.33	-0.00
1081	18	0.13 (-1.95)	0.14 (-1.83)	-2.04	-0.01
1082	12	0.25 (-1.10)	0.19 (-1.47)	-1.63	-0.14
1083	4	0.62 (0.51)	0.55 (0.19)	0.21	-0.05
1084	9	0.25 (-1.10)	0.55 (0.19)	0.21	-1.65
1085	9	0.42 (-0.34)	0.54 (0.17)	0.19	-0.28
1086	5	0.75 (1.10)	0.61 (0.44)	0.49	-0.22
1087	29	0.82 (1.55)	0.73 (1.00)	1.11	-0.71
1089	9	1.00 (5.00)	0.87 (1.89)	2.24	-1.26
1091	31	0.94 (2.71)	0.88 (1.95)	2.58	-0.67
1092	7	0.88 (1.95)	0.88 (2.04)	3.30	-0.00
1097	38	0.85 (1.77)	0.90 (2.20)	4.47	-0.40
1098	9	1.00 (5.00)	0.89 (2.07)	3.54	-1.07
1099	132	0.79 (1.33)	0.89 (2.11)	3.82	-5.59**
1100	36	0.83 (1.61)	0.71 (0.91)	1.01	-1.43
1103	32	0.23 (-1.22)	0.22 (-1.28)	-1.43	-0.00
1104	39	0.20 (-1.37)	0.16 (-1.63)	-1.81	-0.19
1105	11	0.17 (-1.61)	0.18 (-1.50)	-1.67	-0.00

Table 2.8.3 continued

Site	n	p observed	p expected	X	Likelihood
1109	15	0.00 (-5.00)	0.20 (-1.41)	-1.57	-3.28*
1110	40	0.58 (0.31)	0.24 (-1.14)	-1.26	-10.12**
1111	18	1.00 (5.00)	0.95 (3.06)	11.18	-0.82
1112	38	0.85 (1.77)	0.95 (3.06)	11.06	-2.88*
1113	20	0.61 (0.44)	0.52 (0.10)	0.11	-0.28
2012	10	0.46 (-0.14)	0.24 (-1.14)	-1.26	-1.16
2054	14	0.13 (-1.95)	0.26 (-1.03)	-1.14	-0.81
2115	35	0.09 (-2.28)	0.10 (-2.19)	-2.43	-0.01
2116	85	0.08 (-2.43)	0.08 (-2.51)	-4.59	-0.02
2117	65	0.07 (-2.52)	0.16 (-1.64)	-1.82	-2.24*
2118	12	0.81 (1.47)	0.88 (2.00)	2.98	-0.23
2119	42	0.07 (-2.57)	0.08 (-2.47)	-3.66	-0.01
2120	26	0.11 (-2.05)	0.08 (-2.47)	-3.77	-0.21
2121	64	0.02 (-3.89)	0.08 (-2.43)	-2.91	-2.22*
2122	31	0.94 (2.71)	0.93 (2.54)	7.05	-0.03
2124	12	0.81 (1.47)	0.92 (2.50)	6.71	-0.77
2126	46	0.83 (1.57)	0.92 (2.44)	6.30	-2.03*
2127	34	0.97 (3.44)	0.92 (2.42)	6.14	-0.74
2132	16	0.94 (2.71)	0.92 (2.38)	5.86	-0.05
2133	58	0.92 (2.47)	0.91 (2.36)	5.71	-0.02
2134	57	0.89 (2.05)	0.92 (2.40)	5.98	-0.32
2135	46	0.09 (-2.26)	0.08 (-2.43)	-2.82	-0.05
2136	18	0.90 (2.20)	0.92 (2.43)	6.25	-0.04
2138	7	0.88 (1.95)	0.92 (2.47)	6.53	-0.09
2140	28	0.77 (1.22)	0.92 (2.42)	6.10	-2.76*
2141	21	0.85 (1.71)	0.92 (2.40)	6.04	-0.57
2142	11	0.17 (-1.61)	0.08 (-2.48)	-4.03	-0.48
2143	98	0.13 (-1.87)	0.08 (-2.44)	-3.16	-1.66
2144	4	1.00 (5.00)	0.12 (-2.02)	-2.25	-8.58**
2145	20	0.21 (-1.33)	0.08 (-2.43)	-2.85	-1.58
2146	32	0.79 (1.34)	0.45 (-0.18)	-0.21	-7.68**
2147	26	0.45 (-0.21)	0.41 (-0.35)	-0.39	-0.06
2148	18	0.13 (-1.95)	0.08 (-2.44)	-3.11	-0.22
2149	3	0.50 (0.00)	0.07 (-2.56)	-5.50	-1.98
2150	14	0.22 (-1.25)	0.81 (1.43)	1.60	-11.19**
2151	44	0.10 (-2.23)	0.28 (-0.95)	-1.06	-4.40**
2152	109	0.09 (-2.26)	0.08 (-2.44)	-2.96	-0.14
2153	9	0.25 (-1.10)	0.23 (-1.22)	-1.36	-0.01
2154	55	0.76 (1.13)	0.80 (1.41)	1.56	-0.37
2155	23	0.81 (1.47)	0.83 (1.59)	1.77	-0.03
2156	70	0.85 (1.77)	0.64 (0.59)	0.66	-7.81**
2157	14	0.13 (-1.95)	0.52 (0.08)	0.09	-4.86**
2158	31	0.85 (1.73)	0.56 (0.24)	0.27	-6.01**
2159	53	0.13 (-1.87)	0.45 (-0.21)	-0.24	-12.05**
2163	48	0.89 (2.12)	0.92 (2.44)	6.29	-0.21
2164	35	1.00 (5.00)	0.94 (2.79)	8.98	-2.10*
2165	116	0.93 (2.58)	0.94 (2.79)	9.06	-0.17
2166	177	0.07 (-2.60)	0.08 (-2.47)	-3.68	-0.09
2167	36	0.03 (-3.44)	0.08 (-2.50)	-4.37	-0.64
2200	71	0.94 (2.68)	0.92 (2.42)	6.14	-0.16

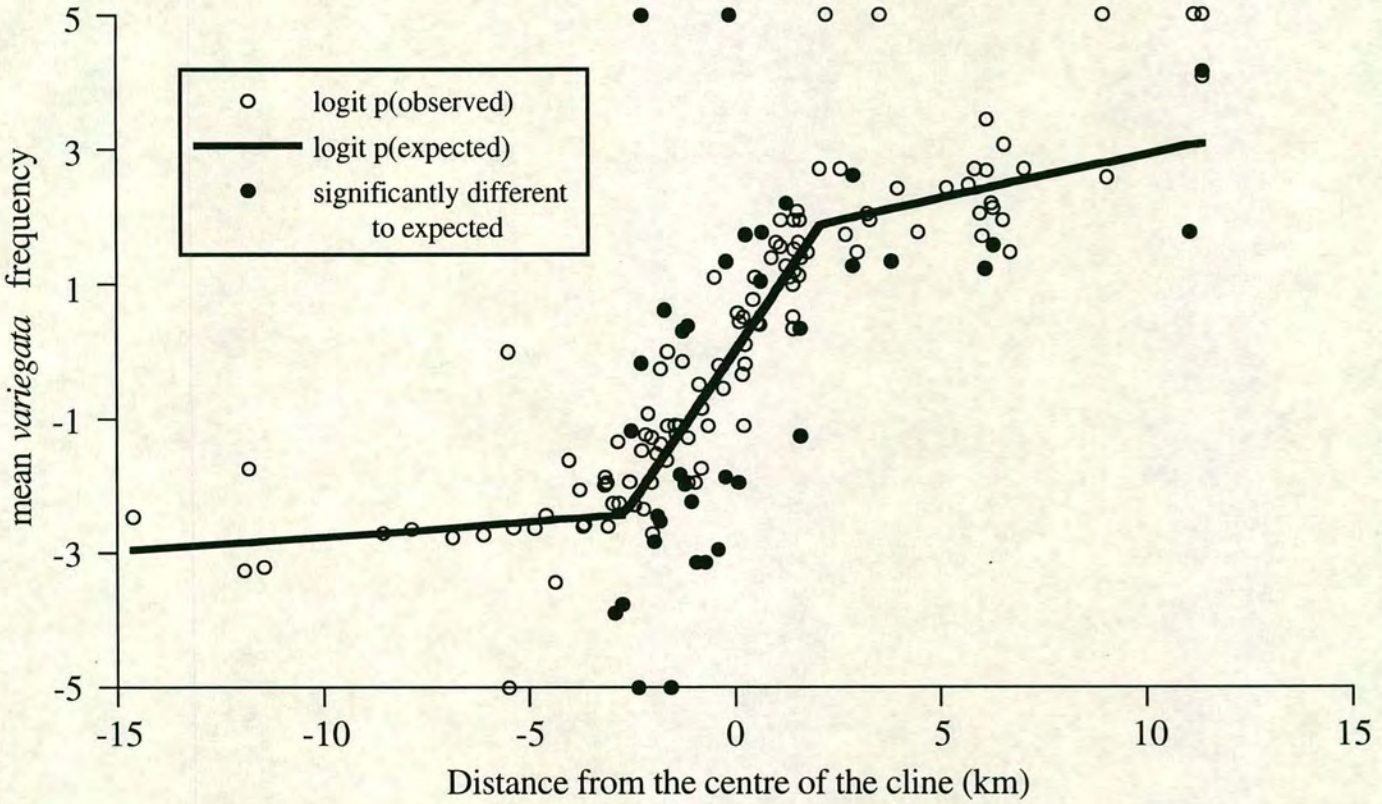


Fig. 2.8.3 A 1-dimensional representation of the most likely cline fitted in two dimensions with a constant width. The likelihood of this cline is L=-251.96. The mean *variegata* gene frequency is logit transformed (=ln[p/q]). Each circle represents the observed gene frequency of one of 147 sites used in the analysis. The expected frequency of a site a certain distance from the centre of the cline is given by the line. This indicates a sharp step in gene frequency in the centre of the cline flanked by two shallower tails representing the rates of introgression on either side of the zone. The sites which deviate significantly from the expected gene frequency are represented by the black circles. Sites where log(p/q) = 5 or -5 are where $\bar{p}$ = 1 and 0 respectively. See text for details (Section 3.8.4).

dependent cline where it is assumed that the distribution of genotypes is largely independent of habitat, selection is against heterozygotes and the cline is stabilised by dispersal and selection. The shape of the cline will be determined by the relative strength of these two forces.

It is already known that the two taxa exist in different habitats. Hybrid zones between *Bombina* are all situated at the transition between uplands and lowlands (Szymura, 1993). Although a tension zone model is a good approximation to the data across the two Polish transects at Cracow and Przemysl, it may be that the habitat transition was smoother there than here. Incorporating a difference in gene frequency according to habitat and/or allowing for variation in width may explain more fully the distribution of genotypes at Pešćenica and help to explain the patterns of disequilibrium and heterozygote deficit outlined above. This will be the aim of the following chapter

Chapter 3

Quantifying a habitat difference between two taxa of *Bombina*

3.1 Introduction

One of the key aspects in any hybrid zone analysis is whether underlying environmental heterogeneity contributes significantly to the position and genetic structure of the zone. Models to analyse hybrid zones fall into two types; those maintained by exogenous selection and those by endogenous selection (Moore and Price 1993, Chapter 1). Exogenous selection, as the name implies, emphasises the maintenance of a hybrid zone through external, ecological variables typified by the geographical selection gradient models of Slatkin (1973, 1975) and Endler (1977). Endogenous selection models explain hybrid zones as a result purely of selection against hybrids (both heterozygotes and recombinants) where fitness is not correlated with any spatial variant; these are referred to as tension zones (Barton and Hewitt 1985, Chapter1). Both types of models can be used to measure the intensity of selection against hybrids but they cannot be used to determine the nature of that selection. It has been argued that most empirical studies of hybrid zones show that they are independent of their environment (Barton and Hewitt 1985). This assumption has been increasingly challenged recently (Harrison and Rand 1989, Moore and Price 1993). This chapter will assess the role of environmental heterogeneity in the *Bombina* hybrid zone at Pešćenica.

Chapter 3 concluded that there is significant residual variation around the model of the stepped cline used to describe the distribution of genotypes . One reason for this is that the two taxa of toads occupy different habitats. The gene frequency of a population may therefore vary depending on the habitat it is found in rather than purely as a function of its position from the centre of the cline.

It has been known since the turn of the century that the two taxa exist in generally different habitats (reviewed in Chapter 1). In general *Bombina bombina* is found in lowland areas in more permanent bodies of water while *B. variegata* exists in more

temporary pools at higher altitudes. The transect at Pešćenica follows this broad trend. In general *bombina* is found in the lowland flood plains around the Sava river to the north while *variegata* occupies the hills to the south (Fig. 2.2.1). Despite the awareness of this habitat difference a detailed analysis of the ecology of the taxa has not been undertaken and there has been no attempt to quantify any differences between the taxa.

The direct question that needs to be answered in the context of this study is whether the gene frequency of a population is in any way dependent upon the habitat in which it occurs? Do different habitats show correlated changes in gene frequency such that populations in a particular habitat type deviate consistently in the same direction from the frequency one would expect were there a homogeneous environment. In order to answer this the following questions need to be addressed:-

1. Are there different habitat types within the study area? More specifically can the range of habitas be reduced to a few general types based on the variables measured at each site.

2. If different habitat types can be identified then is there a relationship between these habitat types and the genotypes of the toads occupying them? Is this relationship consistent across the hybrid zone?

3. Can a correlation between habitat and genotype explain deviations of gene frequency from those expected from the existing model of the cline (Chapter 2), i.e. does the habitat allow one to predict how the observed gene frequency of a population might deviate from that expected in a homogeneous environment?

The aim of this chapter is to identify and quantify any difference in habitat between the taxa and incorporate this into the existing model of the cline. If habitat has a significant effect on the distribution of genotypes across the hybrid zone then allowing for this in the model should reduce the residual variation. The chapter will address each of the above questions in turn. Methods and results will be described where appropriate.

3.2. Are there different habitat types within the study site?

One of the interesting features of this particular study site is the fact that closely neighbouring sites in the centre of the hybrid zone contain populations which show a surprisingly large difference in $\bar{p}$ (Fig. 2.2.1). This might be accounted for by random fluctuations in gene frequency but is unlikely as it is known these animals disperse quite extensively (Szymura and Barton 1986, 1993). It is hard to see how such differences could therefore arise. Szymura classified two pairs of such sites in the 1979 transect and identified each site as a different habitat type. From the outset of the first field season in 1992 it was apparent that different types of toads were occupying different habitat types. Populations could be identified in the field as being *bombina* like or *variegata*-like based on the 'average spot score'. This is a measure of the connections in the pattern of coloration on their bellies, can be scored easily and is highly concordant with genotype (Szymura and Barton 1991 , Chapter 1). In general on the *variegata* side of the zone populations were collected from small puddles formed by wheel ruts which were generally barren of vegetation. On the *bombina* side of the zone populations were sampled from more overgrown ponds or marshy areas.

It did not take long to realise that we could subjectively identify sites as being either *bombina* like or *variegata* like before we sampled the population there. Within the hybrid zone there is a wide range of available habitats. As a test five pairs of nearby sites were subjectively classified as being either more *bombina* or *variegata* like (more pond or puddle like relative to each other). The populations subsequently sampled from them revealed that the site classified as the more *variegata* like always contained the more *variegata* like population of the pair (Table 3.2.1).

Although the analysis was initially based on the spotting pattern of the individuals; the mean gene frequency of the population gave the same result. The probability that this would happen consistently for each pair of sites is 1/32. The strength of this test relies on the fact that the habitat classification was done without any prior knowledge of the kind of toads to be found there. It seems likely therefore, that there is a relationship between the habitat of a site and the mean gene frequency of the population found there. The problem here is to identify differences between habitats objectively.

HABITAT TYPE		
bombina-like	*variegata-like*	Distance (m)
0.21 (1056)	0.59 (1054)	50
0.09 (1045)	0.46 (1046)	20
0.17 (1104)	0.20 (1105)	10
0.13 (1055)	0.25 (1064)	200
0.13 (1043)	0.24 (1044)	300

Table 3.2.1 The *variegata* gene frequency ($\bar{p}$) in neighbouring sites of different habitat. The names of the sites are given in brackets. The distance in metres is given between a pair of sites. Each site of a pair was subjectively identified as being more *variegata*-like or more *bombina* -like. The frequency is always higher in the more *variegata*-like site. If the toads were distributed at random, this pattern would be observed with a probability <0.05.

Methods and materials

Table 3.2.2 lists the variables measured at each site, (Appendix 3.1 gives details for each site). They describe the aquatic site where individuals were actually collected and the terrestrial habitat surrounding that site. The data were recorded when individuals were first sampled. Sites were not consistently re-scored within a season. However the data were retaken if the site was sampled in a subsequent year. The following results refer to sites sampled in 1991 and 1992 only. Sites sampled in 1979 by Szymura have no habitat data associated with them. The data analysis will be in two parts; first the aquatic site will be described and then the terrestrial habitat surrounding it. The relationship between these two habitat types will then be assessed.

The ecological variables measured at each site

The choice of variables measured at each aquatic site was based on the following criteria:-

1.Those known to be important to amphibian habitat (Duellman and Trueb, 1986; Heyer *et al.*, 1994; Laan and Verboom, 1990).

Table 3.2.2. Summary of habitat measures recorded at each site.

A. AQUATIC HABITAT

1. Habitat type (e.g. pond, wheel rut etc.)
2. Width (m)
3. Length (m)
4. Depth (m)
5. Max Bank Depth (m)
6. Bank incline (none, shallow, medium or steep)
7. %Ev (emerged vegetation)
8. %Sv (submerged vegetation)
9. %ShV (shore vegetation in three height classes)
    a) 0-15cm
    b) 15-50cm
    c) >50cm
10. %TC (Tree cover directly overhead)
11. pH-Soil
12. pH-Water
13. Air temperature
14. Water temperature
15. Altitude
16. Pond substrate (e.g. leaf litter, mud etc.)

B. TERRESTRIAL HABITAT

1. Immediate area - general habitat type within 5m radius.
2. Surrounding area - general habitat type within a 500m radius.
3. Region - site situation loosely based on habitat (see section 3.3)

2. Those variables which were not too time consuming to measure. These were mainly physical variables. A large number of sites were covered over the two field seasons. The ecological analysis formed one aspect of the study and so there was a limited time available for it. For this reason no detailed qualitative analysis of the flora or fauna present was undertaken.

There are two ways of reducing such a large number of variables describing the habitat. One would be to cluster sites into distinct types and the other would be to describe a function, based on the variables measured, that accounts for genotype. This latter option is tempting as it is probable that individuals of different genotype do exist in different habitat types. However this would be misleading; genotype cannot be used to distinguish between habitats in general as the gene frequency of a population will not only be determined by habitat but also by its position from the centre of the cline. The same habitat may therefore be occupied by populations of different genotype in different areas of the zone. Genotype may differentiate between habitats but they are likely to emphasise differences either side of the zone rather than differences between sites. An independent measure of habitat is therefore required that distinguishes and classifies sites regardless of the position of that site from the centre of the cline. Only when an independent measure is available can the relationship with genotype be examined.

One of the variables used to describe the habitat, the 'habitat type' distinguishes between sites on the general and subjective basis that they are, puddles, ponds, canals etc. The two main types are puddles and ponds or depressions.

1. Puddles. These are small bodies of water with little or no visible vegetation. They are always generated as wheel ruts by tractors or logging vehicles and are often found in clusters (Fig. 3.2.1a).
2. Ponds. These tend to be larger bodies of water that have more vegetation associated with them; they are often marshy areas (Fig. 3.2.1b). A few are man-made but most appear as "natural" depressions.

There was also a range of sites less easily categorised. These included such sites as drainage and village canals, wells and even sites churned up by the semi-domesticated pigs in the forest (Appendix 3.3.1). It may be that this subjective measure of habitat type based on 'puddles' and 'ponds' demonstrates a genuine habitat difference. A

Fig 3.2.1a Typical 'puddle' sites formed by wheel ruts. Many toads were collected from this habitat type.

Fig 3.2.1b Typical 'Pond' sites where toads were collected. The lower photograh gives an example of a natural depression (see text).

problem however is that many of the sites cannot easily be categorised into a particular habitat type. There are two ways of dealing with this, either the subjective classification is ignored and any difference and clustering of sites is explained independently using the other more objective variables, or, 'habitat type' is used to characterise a difference and the interaction of the continuous variables within these types would define which of the two habitats any unknown sites belong to. The former method is more objective and would in essence allow the data to speak for itself. However information may be lost by not taking the subjective classification into account.

Multivariate methods

There are various ways to deal with multivariate data. The three basic strategies are those of direct gradient analysis, ordination and classification (Gauch, 1982). Direct gradient analysis attempts to measure the distribution of species or samples along a simple environmental gradient e.g. height or depth; ordination aims to represent the relationship between a group of variables by reducing the multidimensional data to a lower dimension, usually two, which maximises the variance of the samples; classification, unlike ordination does not attempt to quantify the relationship between variables but clusters the most similar groups together. Both ordination and classification could be used in this analysis. The aim is to reduce the data to a variable that gives a measure of habitat. This will identify if and how the sites vary. The variable describing habitat could be used alone to look at the relationship with genotype but in order to simplify the data it would be convenient to classify the sites into distinct types. There are two ordination processes that are appropriate; a principal components analysis or a discriminant function analysis. Although ultimately they may give similar results they start out from two view points. The former requires no prior knowledge of how differences between the variables might be associated. It reduces the variables measured to a smaller number of abstract ones with maximal variance. A discriminant function requires that there be at least two known groups prior to the analysis. A linear combination of the variables is then produced which maximises the difference between these two groups relative to the variance within them. This model can then be used to group unknown cases.

The advantage of using a discriminant analysis is that it encompasses both ordination and classification. Therefore it will describe the habitat in terms of a linear combination of variables and classify the sites into distinct types. The subjective

categories, puddles or ponds will be used as the basis for a discriminant analysis. The fact that this is an entirely subjective measurement is no reason to exclude the information. It is reasonable to assume that this measure of aquatic habitat type reflects a genuine difference in habitat; most naive observers would be able to distinguish between wheel ruts and ponds. An argument to exclude this measurement might be that a human perception of a habitat difference is dependent upon such factors as scale and culture. Therefore allowing the variables to cluster independently may reveal more about the relationship of the variables themselves. However the variables chosen may be just as inappropriate in detecting relevant habitat differences. I wish to identify a habitat difference; this can be done subjectively, a discriminant analysis will maximise the differences between two groups based on the variables included in the function and will therefore classify those sites which are not easily categorised into puddle and pond.

For two known groups described by a number of variables (X_1.....X_i) the discriminant function computes a new variable, Z. This is a linear function of the variables such that $Z = a + a_1X_1+a_2X_2+....a_iX_i$; where a_1, a_2....a_i are coefficients estimated from the data. The function maximises the ratio of the between to within groups sum of squares for Z. In order to test whether the two groups defined by the function are statistically different two assumptions are required. Firstly the independent variables must have a multivariate normal distribution and secondly the variance -covariance matrix of the independent variables in each of the two groups must be the same (Dillon and Goldstein, 1984).

One way of testing for multivariate normality is to look at the normality of the variables separately. If the variables do not show normality individually then neither will they be multivariately normal. Clearly the variables to be used in this analysis do not fill this requirement (see below). This means that tests of significance and estimated classification error rates may be biased. Non normality tends to distort individual group error rates such that they are much larger than the optimal value for one group and much smaller for the other. The discriminant function is least affected when the variables are bounded and hence platykurtic e.g. percentages (Dillon and Goldstein, 1984)

Inequality of group dispersion also affects significance testing and classification in discriminant analysis. As the number of variables describing the data increase or when the sample sizes of each group are disproportionate then the significance level

testing for equality of group mean vectors tends to be inflated. This will tend to imply spurious differences between the groups. Therefore the main problem is not whether two groups can be distinguished but whether the two groups are significantly different and whether the model correctly classifies unknown cases.

The sites were divided into two habitat types as far as possible. Group 1 included all ponds and pond-like depressions while group 2 includes all wheel ruts (hereafter known as ponds and puddles respectively). The remaining sites such as canals and wells were undefined. Some sites including Szymura's have no data associated with them at all and have been excluded from the analysis.

In total nineteen variables were measured at each site (Table 3.2.2, appendix 3.1). Of these 16 describe the aquatic habitat type and three the terrestrial. Of the 16 aquatic variables 2 are categorical and the remaining 14 are continuous. A discriminant analysis should be based on as few variables as possible; the more variables that are used the lower the classification error rates but the function is of little practical use. If there were as many variables as sites each site could be classified as a different habitat. Ideally the sample size used in a discriminant function should be the cube of the number of variables (C.Theobald pers.comm.). The sample sizes of puddles and ponds are disproportionate, there are many more puddles identified than there are ponds (55 and 16 respectively). Seven variables have been retained for the analysis; the remaining nine being eliminated on the following criteria:-

1. Maximum bank depth. This was not consistently scored.
2. The pH of the water and surrounding soil. Measurements were taken with pH paper. Unfortunately however there is evidence that pH paper gives unreliable and misleading results (Heyer *et al.*, 1994).
3. Temperature of the air and water. This was only done once at a site and sites were sampled at different times of the day.
4. Tree cover. This can be excluded on the basis that it has more to do with the terrestrial habitat surrounding the site than the nature of the aquatic site itself. An analysis of the terrestrial habitat is carried out separately.
5. Turbidity. This was measured using a Secchi disc. However some sites were too shallow to estimate this accurately. The measure also depended very much on whether a vehicle had recently passed through it .
6. Soil type. The only type ever identified was clay.

7.Length of site. Some sites were so long (e.g. canals) that it was not practical to measure length. These sites were arbitrarily given a length of 100m but cannot usefully be included in the analysis. There is also no obvious reason why length (rather than width) should be a relevant variable.
8. Bank incline and pond substrate. These are both categorical variables. Neither were consistently scored.

These variables will not be considered further. The variables left in the analysis are therefore; width, depth, % emerged and submerged vegetation and % shore vegetation in three height classes. Frequency distributions of the variables show that they are not normally distributed (Fig. 3.2.2).

A stepwise discriminant analysis using both forward and backward selection was applied (Dillon and Goldstein, 1984). Criteria for determining the importance of each variable to the function was based on the minimisation of Wilks' lambda. When variables are considered individually this is the ratio of the within groups sum of squares to the total sum of squares. At each step of the analysis the variable with the smallest lambda is added to the function. The effect that the addition of the new variable has on the function can be estimated with an F statistic. If this is significant at $p = 0.05$ then the variable is retained. Assuming normality this implies that the variable makes a significant contribution to discriminating between the group means. Variables that do not improve the discrimination significantly are eliminated. A detailed account is given by Dillon and Goldstein (1984). The following analysis was done with SPSS version 4.0 for the Macintosh

Results

In total 91 sites were used in the discriminant analysis. Of these 71 were defined within one of two groups by the subjective measure of 'habitat type'(Fig. 3.2.3, Table 3.2.3, individual discriminant scores are given in Appendix 3.1). 16 were assigned to ponds and 55 to puddles. The remaining 20 sites were ungrouped. Overall the percentage of grouped cases correctly classified was 97.18%. Such small error rates are to be expected however with so many variables involved. Despite the inequality of the group covariances (Box's $M = 259$, $df = 10$, $p<0.0000$) the canonical discriminant functions of the two groups are significantly different ($\chi^2_4 = 92.249$, $p<0.0000$ Table 3.2.4.a). The fact that χ^2 is so large suggests that despite the non-normal distributions of the variables and the unequal covariances there is a genuine difference

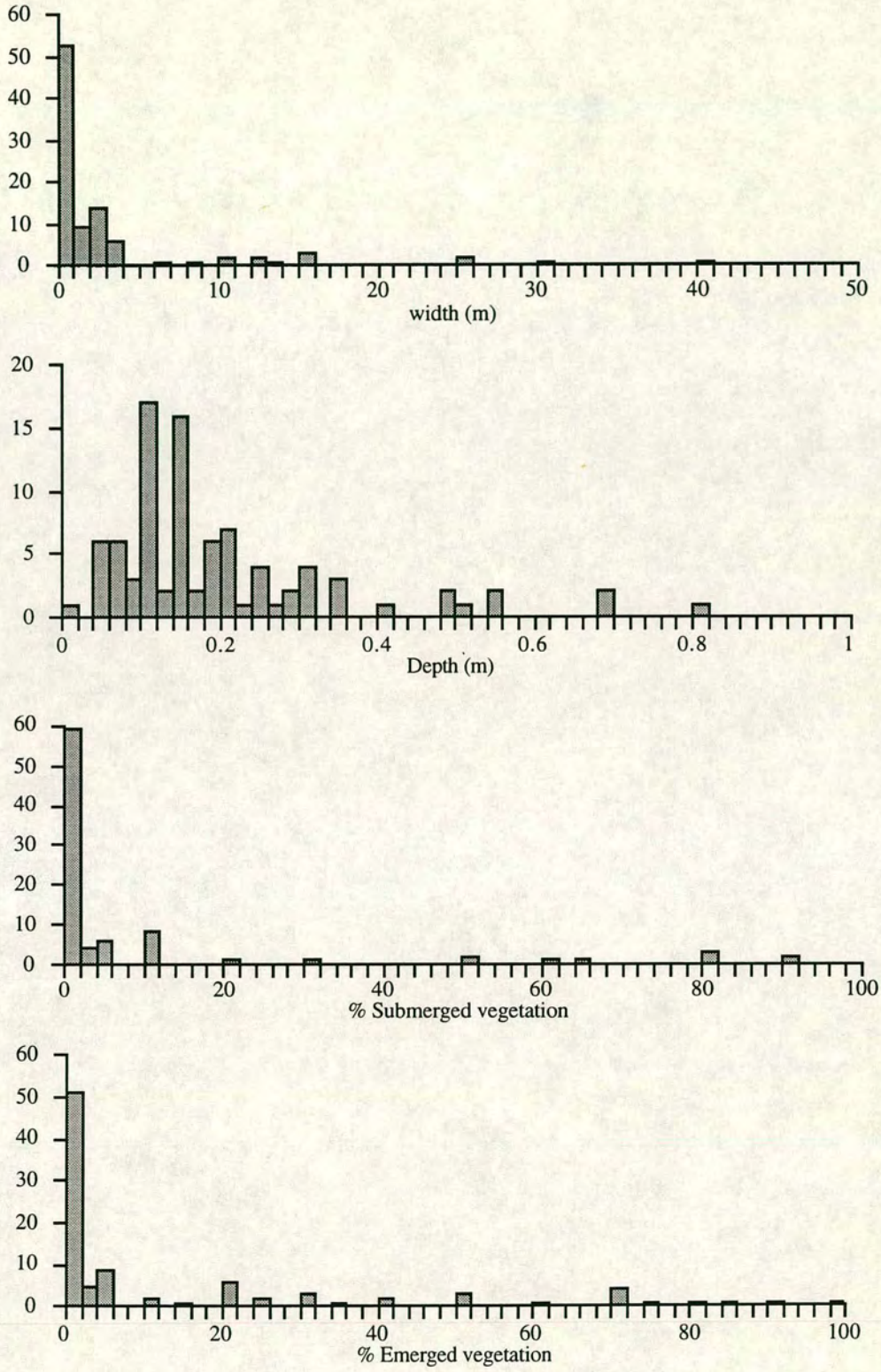


Fig. 3.2.2. Frequency distributions of variables entered into the discriminant function. All seven variables (see over for remaining three) are not normally distributed.

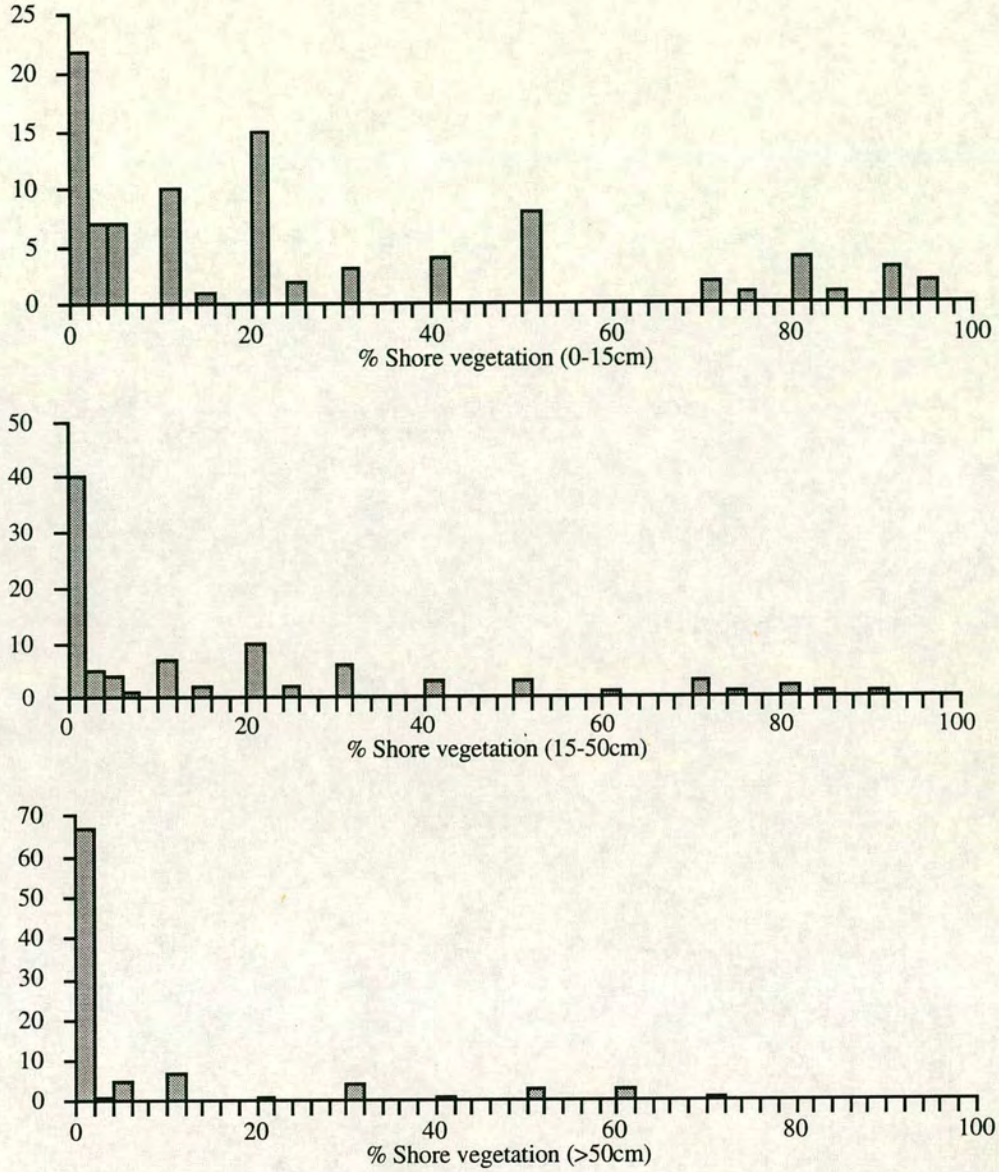


Fig. 3.3.2 continued. Frequency distributions of variables entered into the discriminant analysis.

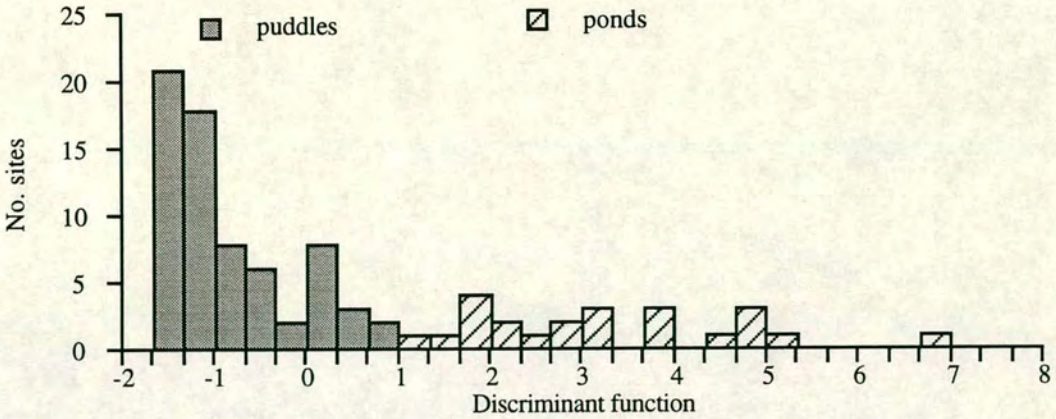


Fig. 3.2.3. Frequency distribution of sites defined by the discriminant function. The mean of the function is -0.898 in puddles (habitat 1) and 3.294 in ponds (habitat 2).

Actual Group	No of sites	Predicted group membership 1(ponds)	2(puddles)
Habitat 1	16	15 93.8%	1 6.2%
Habitat 2	55	1 1.8%	54 98.2%
Undefined	20	7 35.0%	13 65.0%

Percentage of sites classified correctly = 97.18%

Table 3.2.3. Classification results of sites defined by the discriminant function (figure above). Only one site from each of the initially defined groups was misclassified.

Table 3.2.4. Summary results of discriminant analysis.

a. The unstandardised canonical discriminant function coefficients and associated Wilks' lambda. Variables are in the order in which they entered the stepwise analysis

	Overall	Habitat 1(ponds)	2(puddles)	Wilk's λ
% EV	0.046	0.212	0.019	0.342*
Depth	2.434	17.061	6.857	0.295*
Shore V(0-15)	0.018	0.113	0.039	0.256*
Width (m)	0.052	0.196	-0.021	0.247*
Constant	-1.769	-14.018	-1.580	
group mean		3.294	-0.898	

Difference between group means is $\chi^2_4 = 92.25$; $p<0.000$

* $p<0.0000$

b. The mean value of each variable in either habitat, the total mean across both habitats and the correlation of each with the discriminant function.

Variable	Habitat 1 (ponds)	Habitat 2 (puddles)	Total	Correlation with function
%EV	52.50	2.65	3.37	0.80
Width(m)	14.51	0.86	3.79	0.65
Depth(m)	0.44	0.15	0.22	0.43
%SV	49.00	4.02	13.66	0.33
%Shore>50cm	25.13	1.64	6.67	0.27
%Shore0-15cm	45.00	18.62	24.28	0.25
%Shore15-50cm	20.20	11.31	13.21	0.19

between these groups. Of the seven variables initially assigned to the analysis four were used to describe the discriminant function as the other three did not contribute significantly to the between group variance (based on Wilks' lambda). The four variables that describe the difference between puddles and ponds are % emerged vegetation, depth, % shore vegetation (height class 0-15cm) and the width. Most of the variance between habitats can be attributed to differences in the amount of emergent vegetation from a site. This is reflected by the strong correlation of this variable with the function (Table 3.2.4.b).

3.3 Is there a correlation between the aquatic habitat type and genotype?

Gene frequency changes from one side of the zone to the other. One cannot just look at the relationship between the aquatic habitat type defined by the discriminant function and the gene frequency of individuals found there without taking the location of the site into account. One way to do this would be to pool populations a certain distance from the centre of the cline and plot the observed frequency of each population in each habitat as a function of this distance. The position of the cline estimated in the previous chapter could be used. However while this will take the detailed spatial position of the cline into account it will impose certain assumptions, that of a stepped cline for example. However it is important to take geographical information into consideration because of the transition of genotypes across the zone. A crude way to incorporate geographic information is to divide the study area into groups of sites falling into approximately the same region (Fig. 3.3.1). Rough divisions are made on the basis of upland or lowland and arable forested or residential areas. This divides the study site into 13 regions (Table 3.3.1, Appendix 3.1).

The gene frequency of populations in each of the habitat types defined by the discriminant analysis can be plotted as a function of the mean frequency found across both habitat types in each region (Fig. 3.3.2, Table 3.3.2). The populations found in these habitat types vary in gene frequency in the following manner:-

1. The mean gene frequency of populations changes depending on the habitat type such that habitat 2 ('puddles') contain populations with the highest gene frequency (i.e. *variegata* -like) while habitat 1 (ponds) contain populations with the lowest.

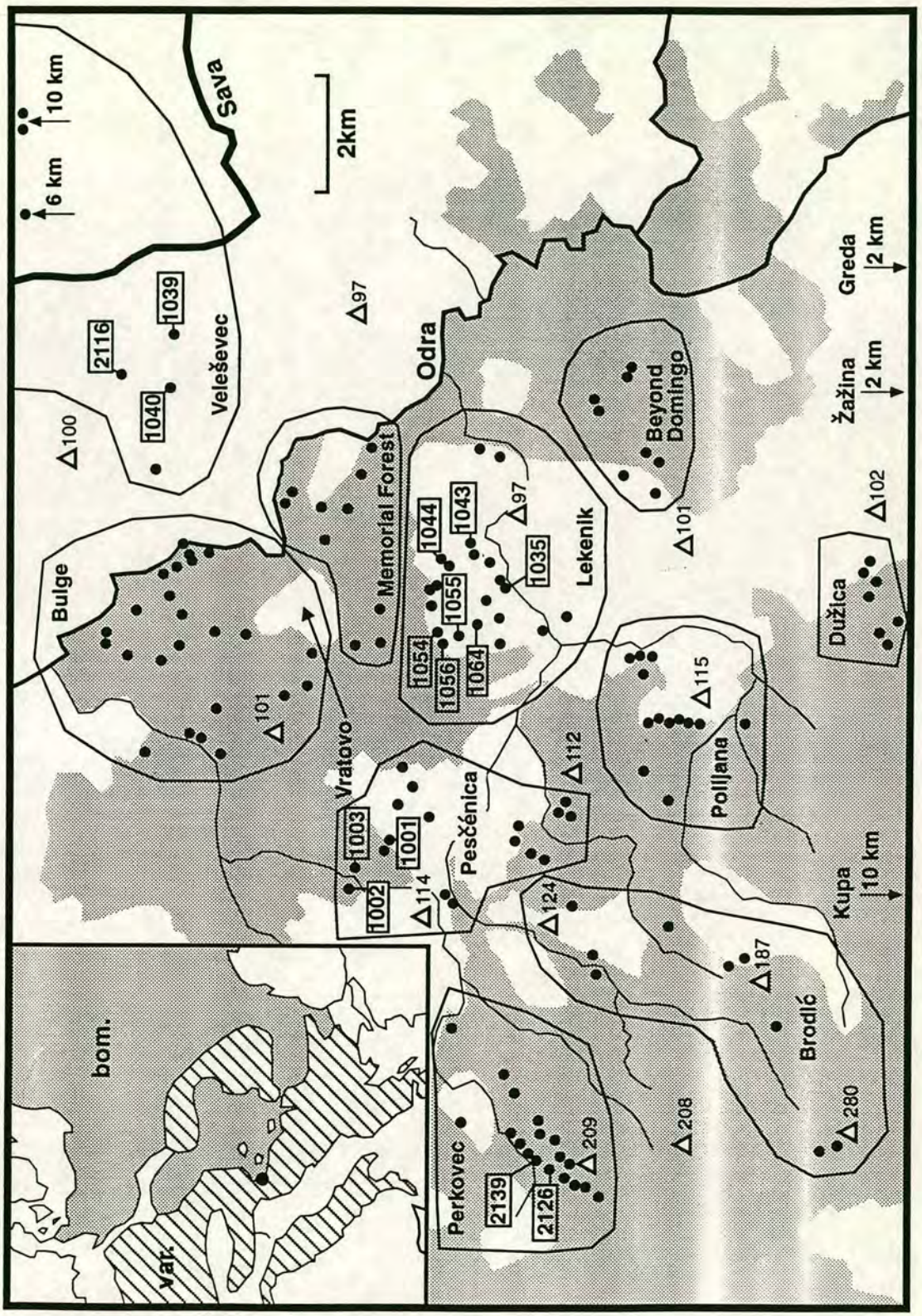


Fig 3.3.1 Map of the Pešćenica transect denoting 'regions' enclosed in circles and sampling locations (•). Integer values next to Δ give altitude in metres. The stippled areas represent forest. The regions are loosely based on similar habitat and altitude. Kupa, Zažina and Greda are all regions outside the range of this map. Sites of particular relevance to this thesis are given in boxes. Recapture experiments (Chapter 4) were carried out at sites marked in Lekenik and Pešćenica. Sites in Perkovec and Veleš eveć (Chapter 5) and the area of Vratovo (marked with arrow) were used in the translocation experiment.

Region	mean $\bar{p}$ of sites	Character	Altitude	Position from Peščenica
Kupa	0.98 (1)	Forest	200	
Perkovec	0.90 (9)	Forest	124-209	west
Brodič	0.89 (3)	Forest	124-280	south west
Bulge	0.35 (16)	Forest	99-101	north east
Memorial.F	0.31 (6)	Forest	98	east
Peščenica	0.82(10)	Residential/arable /forest	103-112	-
Lekenik	0.19(18)	Residential/arable	97-100	east
Polijana	0.72(5)	Arable/forest	100-126	south
Dužica	0.59(6)	Arable/forest	110-127	south
Zažina	0.25(1)	Arable/forest	102	south east
B. Domingo	0.11(6)	Arable/forest	97	south east
Veleševec	0.07(7)	Arable	99-100	north east
Greda	0.05(2)	Arable	100	south south east

Table 3.3.1. Regions dividing the study site, the mean *variegata* frequency of populations within them ($\bar{p}$, and number of sites described in brackets) and the predominant habitat within each (see Fig. 3.3.1 for map).

2. The relationship between habitat type and gene frequency is consistent across all regions of the study site. This eliminates the possibility that the pattern may be due to the fact that wheel ruts are found only on the *variegata* side of the zone and ponds on the *bombina* side.

3. The actual value of gene frequency for each habitat type changes with region. The sites with the highest are in those areas on the *variegata* side of the zone while sites with the lowest gene frequency are on the *Bombina* side. This will reflect the position of the region from the centre of the cline and may also be correlated with the surrounding terrestrial habitat in the area (see below).

Table 3.3.2. The mean frequency of *variegata* alleles for different geographic regions of the hybrid zone. $\bar{p}$ is averaged across both aquatic habitat types, 1 and 2, (mean $\bar{p}$) and separately for each habitat type. Habitat types were defined using a discriminant analysis Habitat 1 = 'ponds' and habitat 2 = 'puddles'. **Fig. 3.3.2.** below shows $\bar{p}$ in each habitat as a function of $\bar{p}$ across both habitats.

Region	mean $\bar{p}$	$\bar{p}$ (habitat 1)	N	$\bar{p}$ (habitat 2)	N
Greda	0.050	0.0312	1	0.0693	1
Veleševeć	0.071	0.0581	6	0.15	1
Beyond Domingo	0.107	0.125	1	0.104	5
Lekenik	0.192	0.122	7	0.237	11
Zažina	0.250	0.250	1	-	-
Memorial Forest	0.307	0.0944	1	0.350	5
Bulge	0.349	0.201	4	0.399	12
Dužica	0.588	0.129	2	0.818	4
Polijana	0.720	-	-	0.72	5
Pešćenica	0.823	-	-	0.823	10
Brodič	0.892	-	-	0.892	3
Perkoveć	0.896	-	-	0.896	9
Kupa	0.984	-	-	0.0984	1

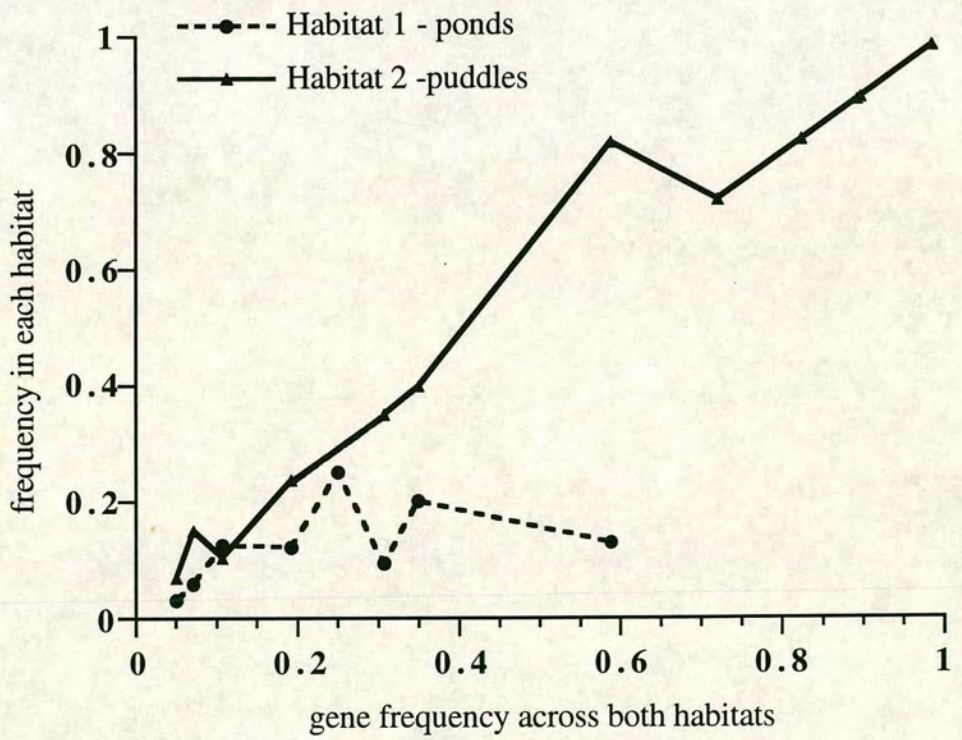


Fig. 3.3.2

4. The difference in gene frequency between populations in different habitats varies with region. Where the gene frequency across all habitat types is small, i.e. on the *bombina* side of the zone the difference between habitats is also small. Where the average gene frequency of the population is 0.5, i.e. in the centre of the zone then the difference in frequency between habitat types is large. Reading directly from Fig. 3.3.2 it is ≈ 0.5. There are few ponds sampled in regions where the average gene frequency is greater than 0.5 so it is unknown what differences between habitats there may be on the extreme *variegata* side of the zone.

What is important is the fact that not only is there a difference between habitat types but that this difference is maintained in all areas where both habitats are found. This would not be expected were genotype distribution random with respect to habitat.

The terrestrial habitat

In general ecological studies of amphibians have centred on a description of the aquatic habitat as this is where most breeding occurs (Duellman and Trueb, 1986; Heyer *et al.*, 1994). The terrestrial habitat however may be equally important in explaining amphibian distributions as it provides corridors of access to the breeding sites and is often the habitat that the amphibian occupies for much of the year.

The habitat surrounding the aquatic sites has been divided into that within a 5m radius (the 'immediate' habitat type) and that within a 500m radius (the 'surround'). Details for each site are given in Appendix 3.1. The surrounding habitat type can be divided into five categories; arable, forest, forest/arable edge, a mixture of vineyard and forest and sites within residential areas. The habitat in the immediate vicinity of the site is more specific. Sites within the forest can be directly beneath the canopy ('forest floor') or in 'clearings' or alternatively are on tracks within the forest which form firebreaks and access routes approximately 6m wide; these are referred to as 'clearways'. Sites outside the forest, apart from two whose boundaries are tarmac, are found in ploughed fields, pasture, or in what I term 'wetland'. Wetland sites are those sites in marshy areas or next to rivers or dykes. They tend to be subject to periodic flooding in wet weather. All this classification is very crude and general. Ideally a detailed analysis of the vegetation should be carried out.

The regions dividing the study site can also be included in this analysis as they are loosely based on habitat as well as their geographic situation in the zone (Table 3.3.1).

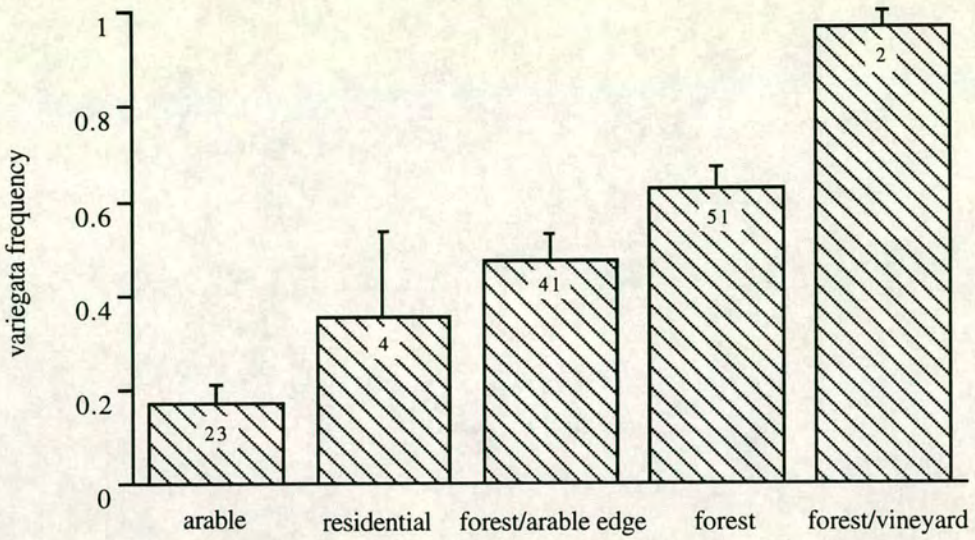
These variables could have been included within the discriminant function. However more information can be obtained by looking at their relationship with genotype independently. For example it may be that the aquatic habitat types defined above exist in particular terrestrial habitats.

1. Gene frequency changes with region across the zone (Table 3.3.1). Those areas with populations of higher gene frequency tend to be in the upland forested regions to the south (such as Perkoveć and Brodi˘č) while those with lower gene frequencies tend to be in the lowlands to the north. This category therefore reflects the general distribution of the two taxa.

2. Gene frequency varies with the "surround" habitat type (Fig. 3.3.3a). Forest populations have higher gene frequencies (i.e. are more *variegata*-like) than forest/arable edge populations which are higher than arable. The two other categories; forest/vineyard mix and residential have very small sample sizes. This pattern does not just reflect the general distribution from upland forest to lowland arable as it is consistent between regions (Fig. 3.3.3b, Table 3.3.3).

3. Gene frequency also changes with the habitat in the immediate area of the site i.e. that habitat within a 5m radius (Fig. 3.3.4). Those habitats associated with the forest (forest floor, clearway and clearing) contain populations of a higher gene frequency than sites such as wetland. However there is variation within the forest such that populations in clearways have a lower mean gene frequency. Wetland contain populations with the lowest gene frequencies whereas sites surrounded by fields, pasture or houses have populations with more intermediate frequencies.

It is difficult to see a consistent relationship between the immediate habitat type around a site and the region as sample sizes are quite small (Table 3.3.4). In general:-
a)Mean gene frequencies in clearways are lower than on forest floors or in clearings apart from the "Bulge".
b) Wetland populations have the lowest gene frequency in all regions (apart from Lekenik where populations sampled from clearways and tarmac have lower frequency however the two sites sampled from the clearways here have a combined population size of three individuals!).



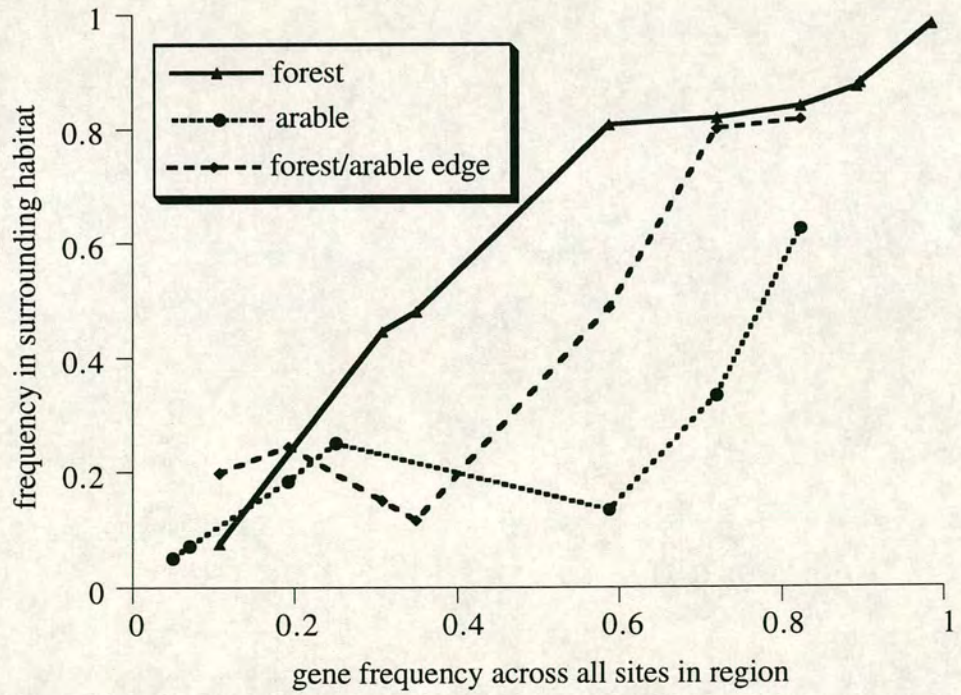


Fig. 3.3.3 a,b. a Gene frequency (and standard errors) of populations in different surrounding habitats i.e. that within a 500m radius of the aquatic site. **b**. The relationship between the gene frequency of sites in the surrounding habitat types compared to the average gene frequency across all sites in a particular region. The relationship of sites in forest, forest/arable edge or arable habitat is maintained.

Region	$\bar{p}$	SURROUND HABITAT arable	forest/ arable	forest	forest/ vineyard	residential
Greda	0.05	0.05(2)	-	-	-	-
Veleševeć	0.07	0.07(7)	-	-	-	-
B.Domingo	0.11	-	0.20(3)	0.08(4)	-	-
Lekenik	0.19	0.18(9)	0.24(9)	-	-	0.18(3)
Zažina	0.25	0.25(1)	-	-	-	-
M.Forest	0.31	-	0.15(2)	0.44(6)	-	-
Bulge	0.35	-	0.12(6)	0.48(16)	-	-
Dužica	0.59	0.13(1)	0.49(2)	0.81(3)	-	-
Polijana	0.72	0.33(2)	0.80(7)	0.82(2)	-	-
Pešćenica	0.82	0.62(1)	0.82(11)	0.84(4)	-	0.88(1)
Brodić	0.89	-	-	0.87(3)	0.96(2)	-
Perkoveć	0.90	-	-	0.88(11)	-	-
Kupa	0.98	-	-	0.98(1)	-	-

Table 3.3.3 Average gene frequency of sites within each region (mean $\bar{p}$) and across different 'surround' habitats within those regions (see also Fig. 3.3.3b). Sample sizes are given in brackets. In general forest sites have a higher gene frequency i.e. are more *variegata* like than arable sites.

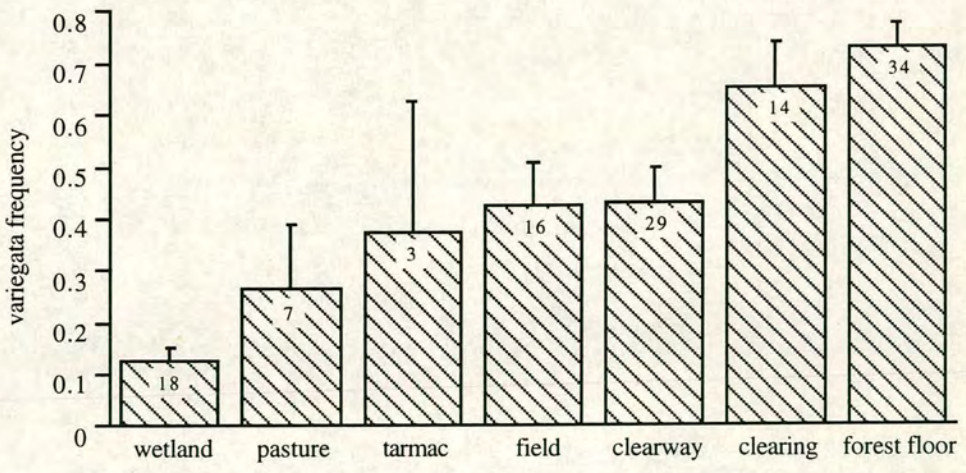


Fig. 3.3.4 Mean gene frequency of populations (averaged across $\bar{p}$) and standard errors in different immediate habitat types i.e. that within a 5m radius of the aquatic site. See text for a definition of the different habitat types.

Region	$\bar{p}$	IMMEDIATE HABITAT						
		clearing	clearway	forest floor	field	houses	pasture	wetland
Greda	0.05	-	-	-	-	-	0.05(2)	-
Vele`ševeć	0.07	-	-	-	-	-	0.09(2)	0.06(5)
B.Domingo	0.11	0.07(3)	-	0.09(1)	-	-	-	0.20(3)
Lekenik	0.19	-	0.03(2)	0.46(1)	0.28(8)	0.12(2)	0.30(1)	0.16(7)
Zažina	0.25	-	-	-	-	-	0.25(1)	-
M.Forest	0.31	-	0.30(4)	0.44(4)	-	-	-	-
Bulge	0.35	0.65(2)	0.39(16)	0.31(2)	-	-	-	0.07(4)
Du`ž ica	0.59	-	0.81(1)	0.81(3)	0.13(1)	-	-	0.12(1)
Polijana	0.72	0.86(4)	0.75(1)	0.74(3)	0.50(3)	-	-	-
Pešcenica	0.82	0.82(4)	0.87(2)	0.82(4)	0.76(4)	0.88(1)	-	-
Brodi`č	0.89	0.88(1)	-	0.89(3)	-	-	1.00(1)	-
Perkoveć	0.90	-	-	0.88(11)	-	-	-	-
Kupa	0.98	-	0.98(1)	-	-	-	-	-

Table 3.3.4 Average gene frequency of sites within each region (mean $\bar{p}$) and across different 'immediate' habitats within those regions (see also Fig. 3.3.4). Sample sizes are given in brackets. The consistency of the relationship between the habitat types within regions is less clear than that shown by the 'surround' habitat.

c) Field, pasture and houses are more variable. They tend to contain populations with gene frequencies intermediate between forest sites and wetland but show no distinguishing pattern. These habitat types do not seem very informative.

The relationship between aquatic habitat type and the terrestrial habitat

The aquatic habitat types classified by the discriminant function are sampled in different proportions from different terrestrial habitat types. In general ponds (habitat type 1) are sampled more frequently from arable areas than forests while the reverse is true for puddles (Fig. 3.3.5). This does not negate the relationship between puddles and ponds as despite this sampling bias, the relationship in the gene frequency between populations from puddles and ponds remains consistent in the different terrestrial habitats (Fig. 3.3.6). Within each habitat the populations from the puddles have the higher *variegata* gene frequency.

It would be tempting to suggest that the difference in the sampling of ponds or puddles from forest or arable habitat reflected the availability of the different aquatic habitats in certain terrain. This may be true; certainly we seemed to come across far fewer ponds in upland forests than in lowland arable areas. However a systematic search of an area would have to be carried out before evidence can be provided for this. The reverse, that there are fewer puddles in arable areas is more easily refuted as these are numerous after rain. However it may be that puddles are more ephemeral in exposed arable areas when not shaded by canopy, effectively reducing their availability.

The alternative explanation for this pattern is that habitat availability is the same everywhere but that habitat occupancy differs between forested and arable areas. In order to test whether occupancy of different habitat types varies across the zone ecology data would have to be collected in sites where no individuals are found and the proportion of different habitat types unoccupied would have to be estimated.

I believe that it is a combination of both factors. In general there is a transition across the zone from upland forest to lowland arable. It is most likely that there are fewer ponds in upland areas and fewer puddles in lowland areas. However there are puddles in the lowland arable land and yet more often than not populations were sampled from ponds. This implies *bombina*-like populations occupy ponds more than puddles.

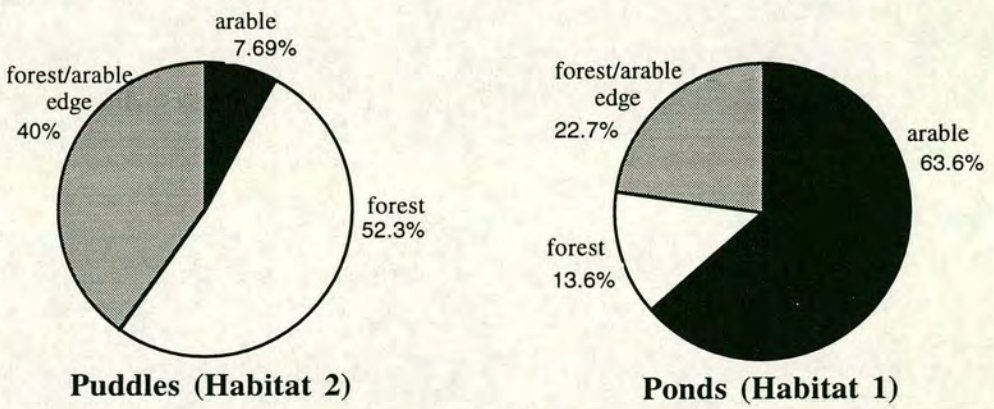


Fig. 3.3.5 The percentage of puddles and ponds in the 'surrounding' habitat type (i.e. within a 500m radius of the site). Only the three major habitat types are included. In general puddles were sampled mostly from forest while ponds were sampled mostly from arable land.

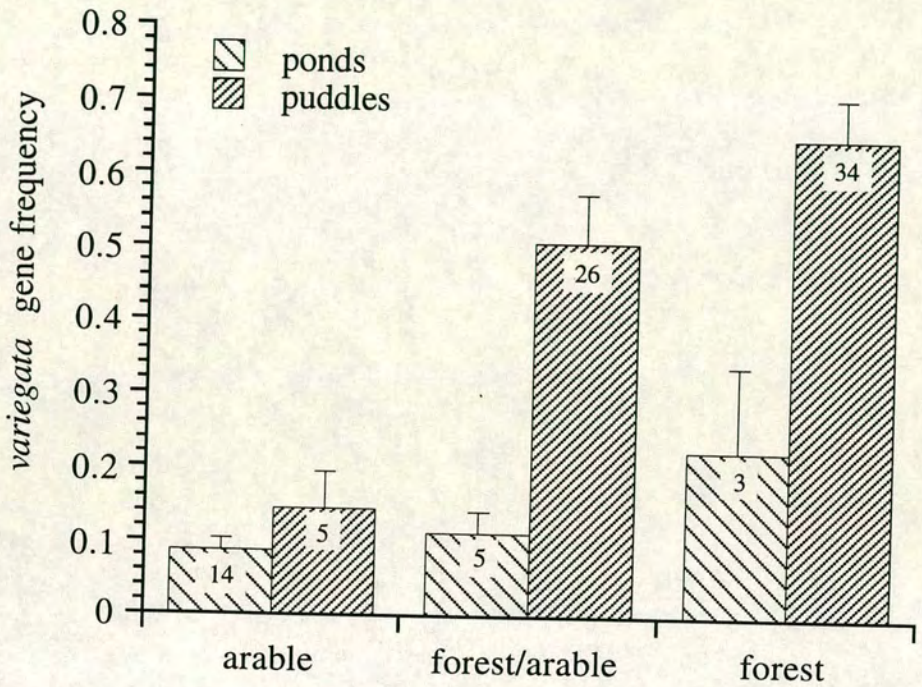


Fig. 3.3.6 The relationship of the mean *variegata* gene frequency ($\bar{p}$) between populations from puddles and ponds in the three main terrrestrial habitat types. The standard error of the mean and the number of sites sampled is given for each bar. In each habitat the populations sampled from puddles have a higher gene frequency than those from ponds.

Comparing different models

It would be ideal to be able to use all the sites that were used in the initial analysis of the cline in Chapter 2 (n = 147). Unfortunately however this is not feasible as the data for the habitat classification are incomplete. The discriminant function above has classified 91 sites as belonging to habitat 1 or 2. There are 26 other sites sampled in 1991 and 92 which have been classified subjectively into either puddles or ponds but have no other ecological measurements associated with them. Taking into account the low classification error rates in the discriminant analysis these sites can be classified subjectively as 1 or 2 with reasonable certainty. However there are still 10 sites from the 91/92 field seasons with no data at all. These will have to be excluded from the analysis. The sites sampled by Szymura have no standardised ecological measurements associated with them. However he described them and assigned some as either *variegata*-like or *bombina*-like. Although it is not certain that his interpretation of the sites is the same, I have classified his sites as far as possible on the basis of his notes (Appendix 3.2). This allows a further 17 sites to be included in the analysis. The initial analysis however will be done without his sites. Once the best fit is found using the 117 sites from 91/92, his data will be incorporated.

The Metropolis algorithm will be used to find the optimum parameter set in the same way as described in Chapter 2. The sites will be assigned to habitats 1 or 2. A number of different hypotheses will be compared. There are two issues to be addressed; first whether allowing for an adjustment in gene frequency according to habitat significantly improves the fit of the model and second whether the cline varies in width along its length. If the width varies then allowing for habitat may reduce or eliminate that variation. If, for example, the cline narrows at the interface between two habitats allowing for the habitat difference would effectively widen the cline at that point. It may be, however, that habitat does not account for all the variation in width. In this case the best fit would require a model where both habitat and width varied.

The differences between these scenarios can be assessed by initially estimating the position and shape of the cline without incorporating either habitat or width. Once an optimum is found this initial parameter set can be the starting point of a subsequent run where width and/ or habitat is included. Therefore for a number of replicate trials four different models will be assessed.

1. Constant width along the cline and no differences between habitats.

2. Constant width along the cline and allowing for a difference in gene frequency according to habitat i.e. α is allowed to vary between Metropolis runs.

3. Variable width along the cline but no habitat difference.

4. Variable width along the cline and α.

The models 2,3 and 4 are not independent as their initial starting points are set from the results of model 1.

Based on the results from the previous chapter the cline will be described by nine segments each 4km long with smoothing superimposed. The cline will be initially determined in two dimensions and then reduced to one dimension to allow for a detailed analysis of cline shape.

Limits on all the parameters included in the model can be estimated by setting the temperature in the Metropolis algorithm to 1. This generates a distribution of the parameters whose density is proportional to the likelihood. Replicate runs are made and the distributions of each parameter can be plotted graphically. Appropriate 95% confidence limits are contained within the area of the graph bounded by $\pm$ 2.5%.

The most likely two dimensional cline

Twelve trials were completed for each model (Table 3.4.1). The cline is best described by a model which allows for both variable width along the cline and a difference in habitat. The likelihood of this cline is logL = -114.60 with 93 residual degrees of freedom (ΔL_{10} between models 1 and 4 = 57.22; $p<0.001$). Although the starting point of each trial varies this model gives consistently significantly better results than the others in all trials (Fig. 3.4.2). Allowing for habitat or varying width separately improves the fit of the basic model (ΔL_1 = 22.16, $p<0.001$ and ΔL_9 = 27.88, $p<0.001$ respectively) and there is no significant difference between the two. Given the analogy of an adaptive landscape this implies that each of these two models rests on a different but similar sized peak. However the highest peak found is when both variables are incorporated.

Table 3.4.1

Trial	Model: Constant width no habitat 1	Constant width with habitat 2	Varied width no habitat 3	Varied width with habitat 4
1	-143.21	-132.13	-123.56	**-114.60**
2	-145.15	-132.50	-123.64	-115.70
3	-165.15	-147.86	-150.41	-128.76
4	-144.39	-132.72	-124.60	-119.83
5	-166.91	-154.15	-136.26	-130.32
6	-156.81	-141.42	-124.92	-121.76
7	-165.75	-155.07	-143.71	-132.94
8	-167.80	-149.49	-131.43	-130.71
9	-154.37	-137.59	-133.18	-127.12
10	-157.22	-144.08	-129.94	-124.16
11	-152.41	-136.34	-128.74	-122.12
12	-158.28	-141.49	-137.96	-124.06

b. Difference in 2xLikelihood between best results.
degrees of freedom and χ^2 given in brackets for $p<0.001$.

	Constant width	Constant width + hab	Varied width
Constant width + hab	22.16 (1, 10.83)		
Varied width	39.30 (9, 27.88)	17.14 (9, NS)	
Varied width +hab	57.22 (10, 29.59)	35.06 (9, 27.88)	17.92 (1, 10.83)

Table 3.4.1. a. Likelihoods of 12 trials for four different models. The mostly likely cline is $logL_{93} = -114.60$ which allows for habitat and varied width (Trial 1 in bold). b. The difference in log likelihood between the highest likelihood for each model (see Fig. 3.4.2)

Incorporating the 17 sites with a classified habitat sampled by Szymura gives a likelihood of $logL_{110} = -145.59$. This is done by reading in the parameter values defined for the most likely cline (from Trial 1 where logL = -114.60) but using the increased data set. The Metropolis algorithm is then allowed to move uphill to the best value. This results in an initial likelihood of logL = -160.18. The algorithm is then set to a temperature of 1 to generate a distribution around this cline proportional to the likelihood. This not only generates limits around the best cline but, because setting the temperature to one allows very small changes to the parameters over many replicates the fit of the cline improves resulting in a final likelihood of -145.59. It is difficult to compare directly whether the addition of these sites improves the fit of the model or not. It is worth including them as they have large sample sizes. Values of the distance of each site from the centre of the cline, the width of the cline at that point and the likelihood of the observed gene frequency given the model are presented in Appendix 3.3A.

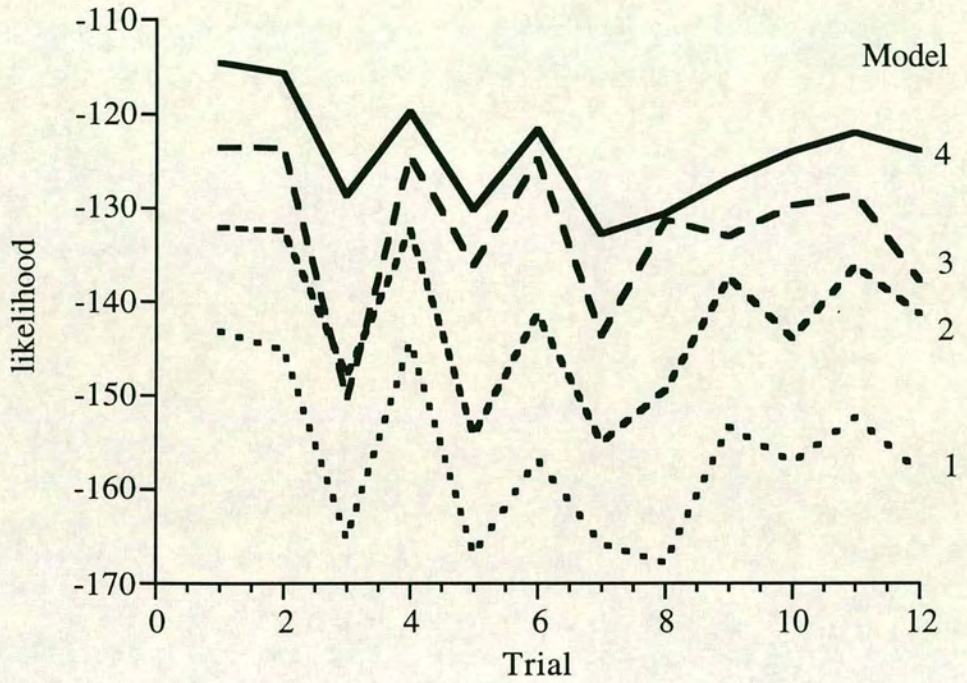


Fig. 3.4.2 Graphic representation of the results of comparing the likelihood of four models (see Table 3.4.1, previous page). Twelve trials were completed. The graph shows that over all trials model 4, which allows for habitat and varying width, is consistently more likely than the other models (see text for details).

The shape and course of the cline

The detailed analysis of cline shape will be discussed below when it is reduced to one dimension. At this stage it is therefore only relevant to discuss the width of the cline and α (the estimated difference in gene frequency according to habitat). Other parameters describing the cline (barriers to gene flow, rates of introgression etc. see Chapters 1 and 2 for descriptions) are given in Table 3.4.2 but will not be considered here.

Variation in width

The width of the cline varies considerably along its length (Table 3.4.2 , Fig. 3.4.3a, Fig. 3.4.3b shows cline on the two dimensional map). It ranges from 6.73km at the beginning of segment 2 to almost zero at the beginning of segment 9. The difference in gene frequency between habitat types is estimated as $\alpha = 0.23$ (0.17-0.37). This means that at the centre of the cline populations in different habitats will differ in gene frequency by 0.115; habitat 2 (puddles), will contain populations with higher gene frequency. This estimate is much lower than the observed difference of 0.5 between the habitat types, (Fig. 3.3.2), calculated in the last section. Possible reasons for this will be explained later.

Limits for all of the parameters were generated by setting the Metropolis algorithm to a temperature of 1 for 5980 replicates. The limits for the barriers to gene flow, rates of introgression and alpha will be looked at in detail below when the cline is reduced to one dimension. The limits around all the values for the widths and angles describing the course of the cline are highly variable (Table 3.4.2). This reflects the amount of data determining the parameters at each segment. There will always be a problem when data collection is not even that limits will vary between segments. One segment has no data associated with it (segment 8). There is no reason however to doubt that the variation in width is real as the likelihood of the cline significantly improves when the width is varied. However it can be argued that where a segment is only described by a few sites (e.g. segment nine), the reason the cline narrows might be due to the close juxtaposition of two habitats. A narrowing of the cline will significantly improve the fit here, but increased sampling might reveal a wider scatter of habitats. The true width of the cline would therefore be greater than that estimated. However this can only be confirmed by more sampling. Given the significant association between habitat type and genotype it is reasonable to assume that the width of the cline

Table 3.4.2. Values and limits (in brackets) of the most likely cline fitted in two dimensions with variable width and α(L_{98} = -145.59). Parameters were varied over 5980 replicates. Each replicate is one step in a random walk which has a probability density equal to the likelihood. Only those sites with a habitat classification have been included. Angle 9 and width 10 are spurious as there is no following segment to define them.

Width and angles

segment	width (min, max)	angle (min, max)
1	5.34 (0.45, 12.69)	0.09 (-0.22, 0.20)
2	6.73 (0.50, 10.71)	-0.76 (-1.01,-0.15)
3	1.77 (0.44, 3.35)	-1.19 (-1.60,-0.97)
4	1.68 (0.83, 3.16)	0.55 (0.29, 0.71)
5	3.41 (3.08, 5.82)	4.55 (4.33, 4.57)
6	3.99 (2.15, 7.08)	-1.46 (-1.53,-1.05)
7	1.80 (0.71, 3.42)	-0.25 (-1.29,-0.10)
8	1.34 (0.09, 4.07)	4.01 (3.13, 4.35)
9	0.06 (0.01, 4.03)	254.93 (-0.39,294.67)
10	3.02 (0.22, 8.80)	-

Other parameters (see section 3.4; Chapter 2.8.2 gives definition of variables)

B_b/w (km)	18.20 (5.39, 368.54)
θ_b	0.01 (0.00, 0.06)
B_v/w(km)	1.30 (0.52, 3.47)
θ_v	0.16 (0.06, 0.63)
α	0.23 (0.17, 0.37)

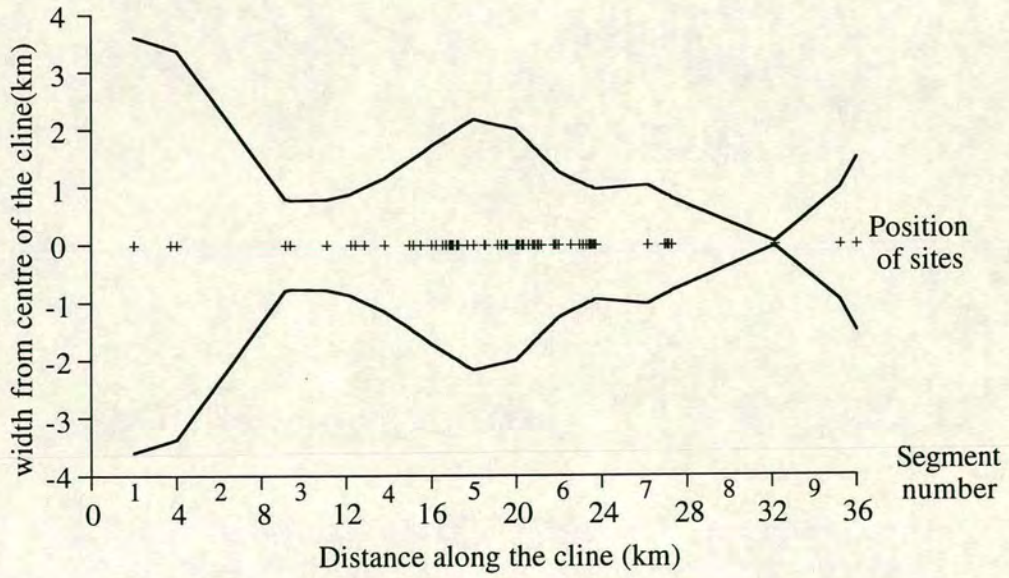


Fig. 3.4.3a The variation in width along the cline (see Table 3.4.2 above). The cline is described by 9 segments each 4km in length. The positions of sites along the cline show the variation in the amount of data describing each segment.

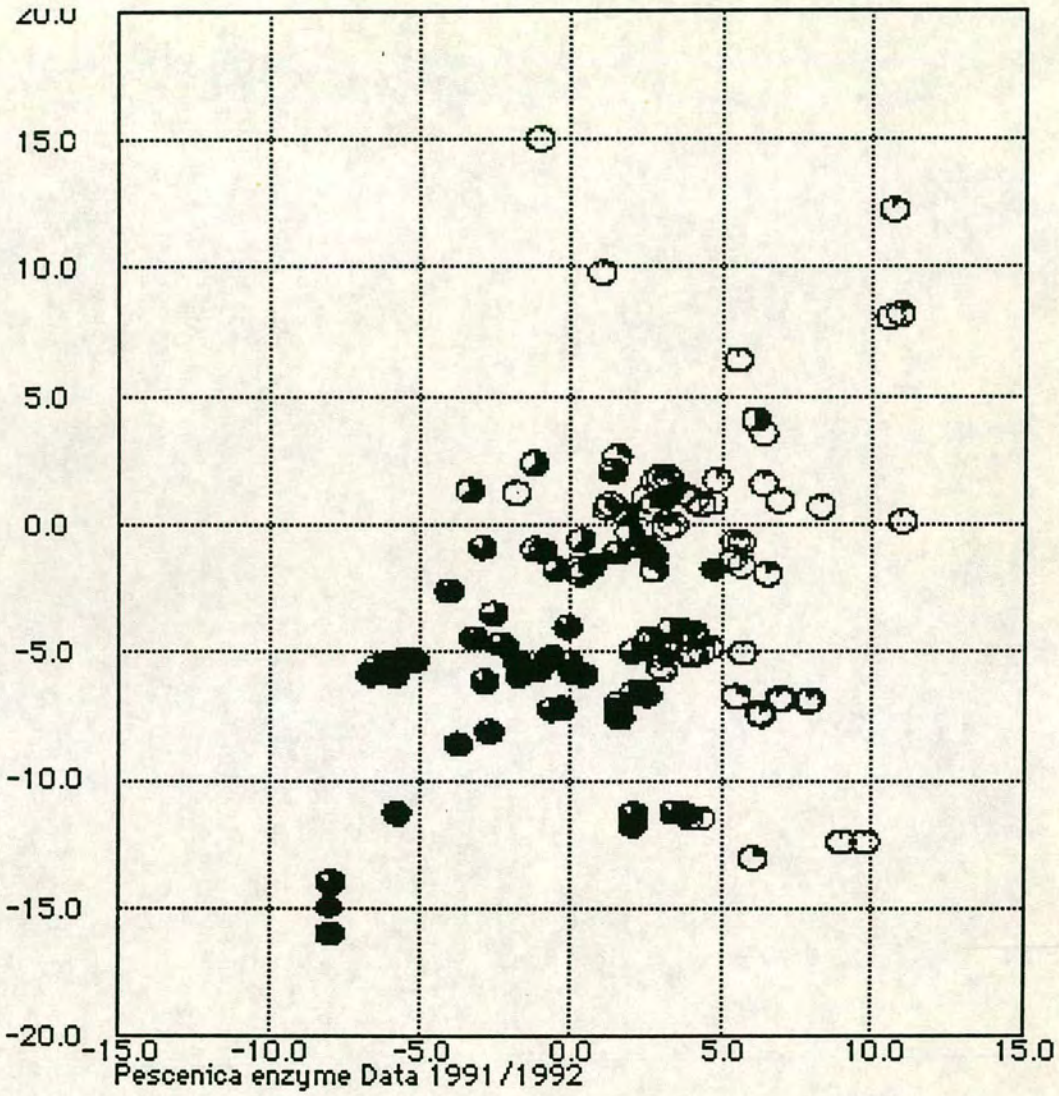


Fig. 3.4.3 b The fitted cline in two dimensions. This allows for variable width and a difference in gene frequency between habitats (see text for details). The centre of the cline is marked in red and the width in green. The location and mean *variegata* gene frequency ($\bar{p}$), of populations are represented by pies. Axes are measured relative to the global origin in km.

will vary. The cline is wider for segments five and six which co-incide with the regions called the 'bulge' and 'Lekenik'. The former is an area of lowland forest, the latter is mixed arable with forest to its north. Both regions have a wide variety of habitats where a mixture of genotypes co-exist. Segment 9 goes through the region called Du`ź ica. Here there is a sharp transition from upland forest to lowland arable; one would expect the cline to be narrower here.

Constraining the width to be constant

In order to estimate dispersal rates and the strength of selection maintaining the barriers to gene flow (Chapter 1, Chapter 6), a measure of the average width along the cline is required. Allowing for a difference in habitat but constraining the width to be constant results in a cline where $\log L_{102}$ = - 132.13 (Trial 1, Table 3.4.1). If Szymura's sites are included and limits are determined over 7201 replicate runs as described above then the most likely cline is $\log L_{119}$ = -176.14 with a constant width of 4.67km. The limits are wide (2.32-5.13; Fig. 3.4.4) but do not encompass the extremes of width seen when this parameter is allowed to vary. The lower limit is especially wide due to a long narrow tail. This is not an average width along the cline, it is the most likely width where the likelihoods are summed across all segments. However this is still an approximation as varying width significantly improves the fit (ΔL_9 = 30.55; $p<0.001$).

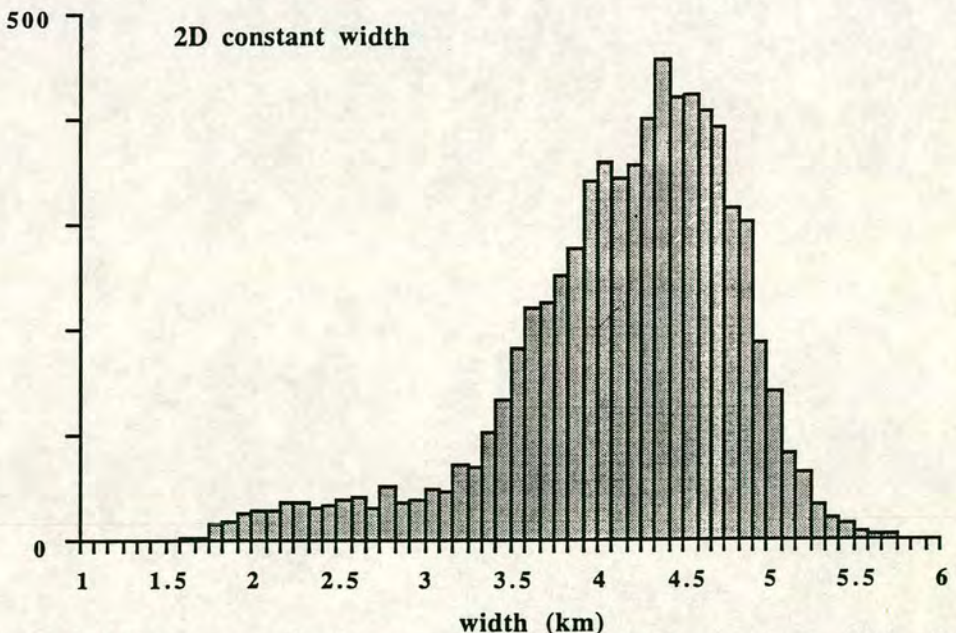


Fig. 3.4.4 Frequency distributions of the most likely constant width of the cline (with α) The Metropolis algorithm was set at T = 1 for 7201 replicate runs creating a distribution whose density is equal to the likelihood. Confidence limits are defined within the area of the graph bounded by 2.5% tails. See section 3.4 for details.

Reducing the two dimensional cline to one dimension

To look in detail at the other parameters describing the model the best two dimensional cline can be reduced to one dimension by plotting the gene frequency of a site relative to its distance from the centre of the cline. As the cline varies in width along its length the distance of each site from the centre is standardised by the width of the cline at that point. This produces a dimensionless co-ordinate. The co-ordinate is estimated from the most likely cline above where width and α are allowed to vary ($\log L_{110}$ = -145.59, Appendix 3.3A). A one dimensional cline means that the number of parameters allowed to vary at any one time are reduced (it eliminates segments and angles). This allows the parameters such as the barrier strength, rate of introgression and alpha to be examined in more detail.

The shape and course of the cline when the variance is estimated from the discordance between loci (i.e. $F_{st}' = 0.0068$)

In Chapter 2 the variance of the mean gene frequency was estimated from the discordance between loci. This gave a value of $F_{st}' = 0.0068$ (Chapter 2 section 5). This was used in conjunction with disequilibrium to account for sampling error and from it the effective sample size for each site was estimated. However this variance does not take into account concordant fluctuations between sites over and above that expected from disequilibrium. This can be estimated directly from the data by allowing F_{st} to be varied alongside the other parameters. This measure of F_{st} is estimated in the normal way as and is therefore different from that of F_{st}'. Therefore the following results will be divided into two. First the parameters describing the shape of the cline will be estimated given an F_{st}' of 0.0068. Then the most likely value of F_{st} will be estimated alongside the other parameters describing the cline.

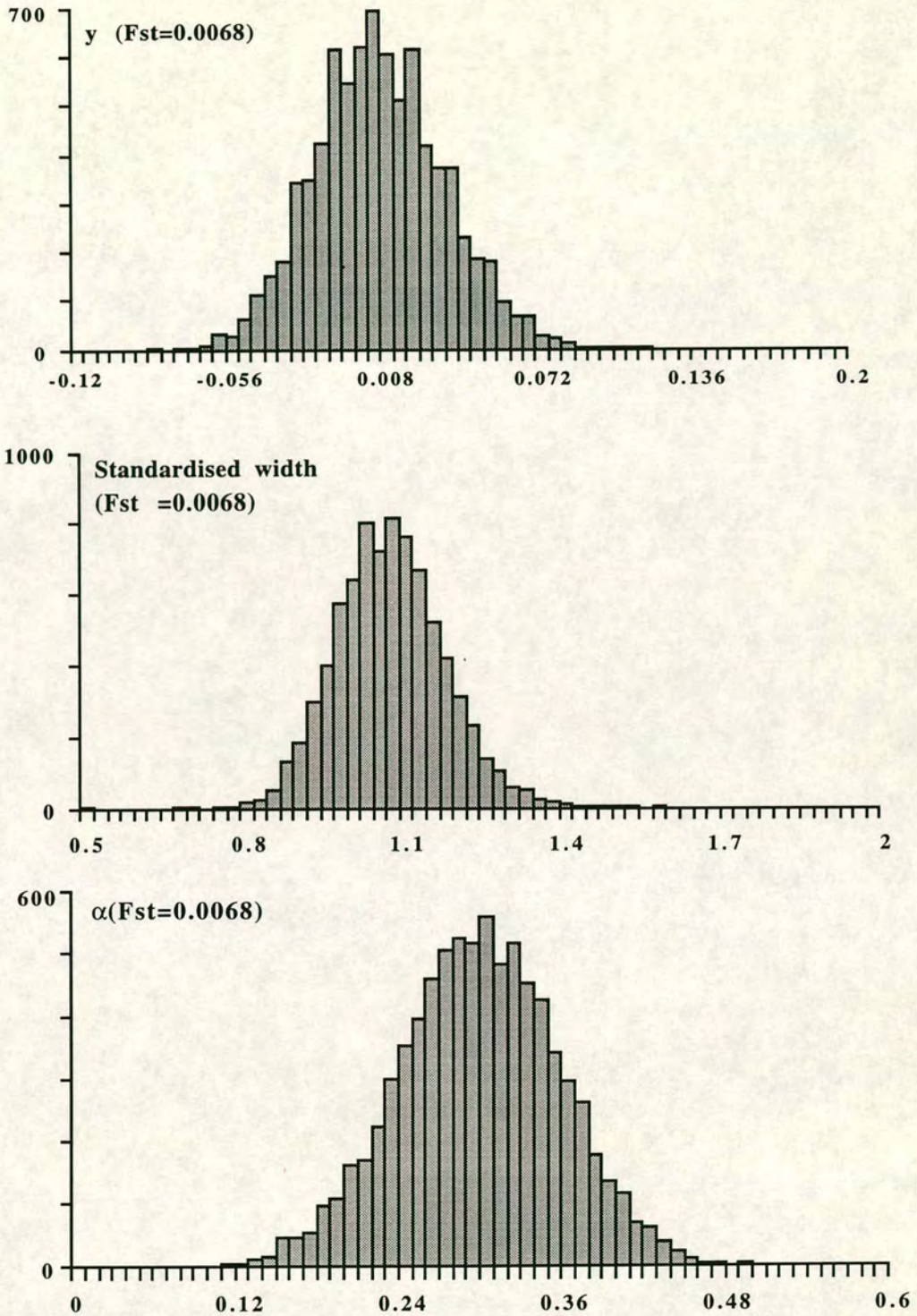


Fig. 3.4.5 Frequency distributions of the most likely estimates for the position of the centre of the cline (y), the standardised width (w) and the difference in gene frequency according to habitat (α). The cline is analysed in one dimension. The Metropolis algorithm was set at T = 1 for 8000 replicates creating a distribution whose density is equal to the likelihood. The variance in gene frequency around this cline was set at $F_{st}' = 0.0068$ (determined from the discordance between loci). Limits on these estimates are defined within the area of the graph bounded by 2.5% tails (see text and Table 3.4.3).

The initial starting point for the parameters describing the shape of the cline are those which are defined for the best two dimensional cline (including Szymura's 17 sites, Table 3.4.2). The Metropolis algorithm is set to 1 allowing a random walk through parameter space. This will give the most likely value for each parameter and a distribution around it whose density is proportional to the likelihood. Given an F_{st}' of 0.0068 the values and limits of the parameters describing the shape of the cline were estimated over 8000 replicate runs (Table 3.4.3).

The width and position of the cline

The position of the cline relative to the predicted centre (y) and the standardised width are estimated accurately. Their limits are narrow (Fig. 3.4.5, Table 3.4.3) and plotted against each other they show a reasonably tight distribution (Fig. 3.4.6). The deviation of the cline depending on habitat is $\alpha = 0.30$ (limits 0.18, 0.41). An α of 0.3 means that in the centre of the cline the mean gene frequency of populations in either habitat type differ by 0.15.

The barriers to gene flow and rates of introgression

As α is constrained to be the same across the cline the barriers to gene flow and rates of introgression in either habitat type on the same side of the zone will not differ. Ideally one would assess the shape of the cline independently in each habitat. This will not be done here. Therefore estimates are combined over both habitat types. The rates of introgression of *Bombina* genes into the *variegata* gene pool are estimated as 10 times that of the reverse ($\theta_v = 0.098$, $\theta_b = 0.001$; Table3.4.4). However the limits to the rate of decay of introgressing alleles into the *bombina* side of the zone are extremely wide (θ_b; 0-0.054) and overlap with estimates for the limits of θ_v (0.045-0.190, Table 3.4.3 Fig. 3.4.7). The limits on the *bombina* side of the zone are wider as there are fewer samples on this side of the zone.

The barriers to gene flow are also very different. The barrier estimates given in the following graphs have all been measured relative to the width of the cline. The distance from the centre of the cline has been standardised by the width. Assuming that the width of the cline is approximately 4.67km (when constrained to be constant) then the actual barrier strength is obtained when the values from the graph are multiplied by 4.67. The graphs are to show the distributions of the most likely estimates and will not be converted. The barrier to gene flow reflects the step in gene

The shape of the cline in one dimension.		
Parameter	**Maximum likelihood estimates**	
F_{st}	0.0068 (constrained)	0.025 (0.015, 0.046)
y	0.06 (-0.04, 0.06)	0.015 (-0.05, 0.09)
standardised width	1.05 (0.090, 1.29)	1.08 (0.88, 1.39)
Bb/w	23.46 (7.30, 397.33)	40.85 (6.83, 666.20)
Bv/w	2.53 (1.45, 5.18)	3.33 (1.46, 8.07)
θb	0.010 (0, 0.054)	0.004 (0, 0.056)
θv	0.098 (0.045, 0.190)	0.071 (0.026, 0.188)
α	0.30 (0.18, 0.41)	0.31 (0.13, 0.42)
No of replicates over which estimate was determined.	8000	3563
$\log_e$ Likelihood	-157.12	-118.49

Table 3.4.3 The most likely estimates and limits to the parameters describing the shape of the cline when the variance in gene frequency is constrained to 0.0068 (F_{st}') or allowed to vary (F_{st}). Estimates were determined over a number of replicate runs with the Metropolis algorithm set at T = 1. This generates a distribution around the maximum likelihood estimate with a density proportional to the likelihood. Note the wider limits to all the parameters when F_{st} is estimated as 0.025.

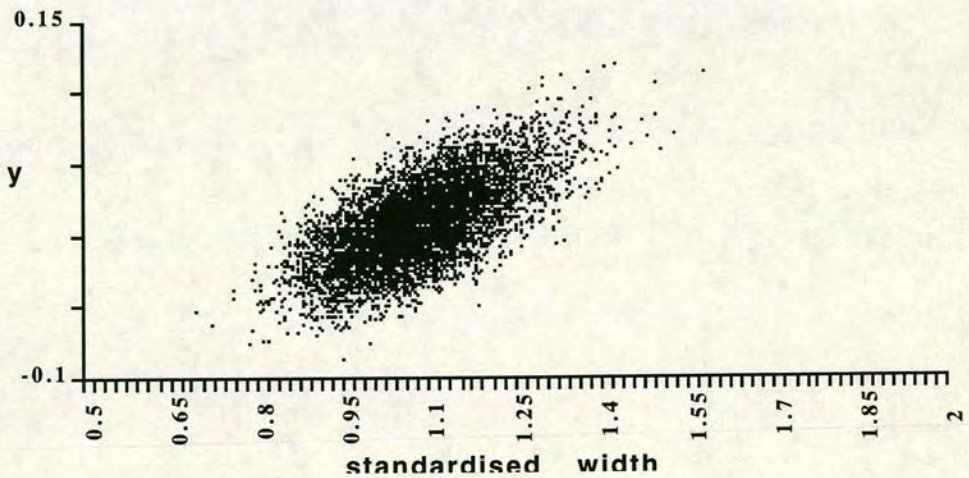


Fig. 3.4.6 The distribution of the position of the centre of the cline (y) plotted against the standardised width given a variance in gene frequency of $F_{st}' = 0.0068$. Each point represents one replicate of 8000 runs with the Metropolis algorithm set to T = 1. This generates a distribution whose density is proportional to the likelihood.

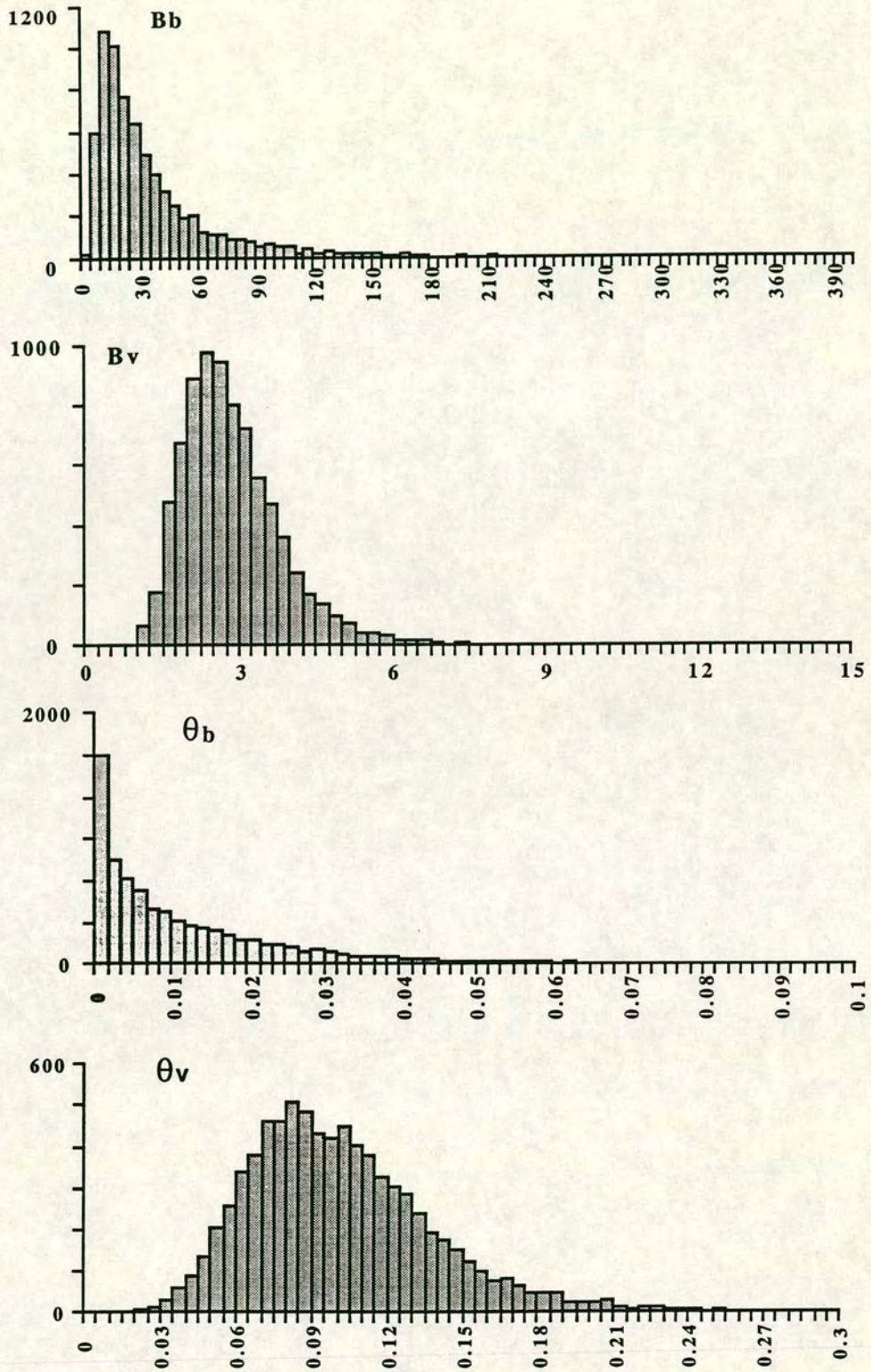


Fig. 3.4.7 Frequency distributions of the barriers to gene flow (Bb, Bv measured in widths, see text) and rates of introgression (θb, θv) either side of the zone. The Metropolis algorithm was set at T = 1 for 8000 replicate runs creating a distribution whose density is equal to the likelihood. These limits are determined when the variance in gene frequency is estimated as $F_{st}' = 0.0068$ See section 3.4 for details.

frequency in the centre of the cline. Although inferences can be drawn from this regarding the strength of the barrier that would generate the step it has to be remembered that it is a description of the shape of the cline in the centre (Chapter 1). It can be regarded as the equivalent distance of unimpeded habitat a neutral allele would have to travel.

The barrier to *variegata* alleles on the *bombina* side of the zone is equivalent to 109.56km (Bb/w = 23.46; B = 23.46 multiplied by 4.67) while the barrier to *bombina* alleles on the *variegata* side of the zone is only 11.81km (2.53 multiplied by 4.67km). However one has to accept all these values with extreme caution. Like the introgression rate, on the *bombina* side of the zone the limits on the estimate of the barrier strength is very wide (Table 3.4.3, Fig. 3.4.7). Again this reflects the lack of samples on that side. The limits on the *variegata* side where there are more data are much narrower. However unlike the rates of introgression the limits to the estimates of barrier strength on either side of the zone do not overlap. This implies an asymmetric barrier to gene flow; i.e. the barrier to *variegata* alleles introgressing on the *bombina* side of the zone is greater than the barrier to *bombina* alleles introgressing on the *variegata* side. Put another way the step in allele frequency on the *bombina* side of the zone is more marked than that on the *variegata* side.

These results have to be treated with caution. It is difficult to dissociate the relative contributions of barrier strength and introgression to the shape of the tails (Fig. 3.4.8). Two scenarios are equally plausible. Either there is a strong barrier and weak introgression or *vice versa*. Also, although the limits to the barrier strengths either side of the zone do not overlap, the gap is small.

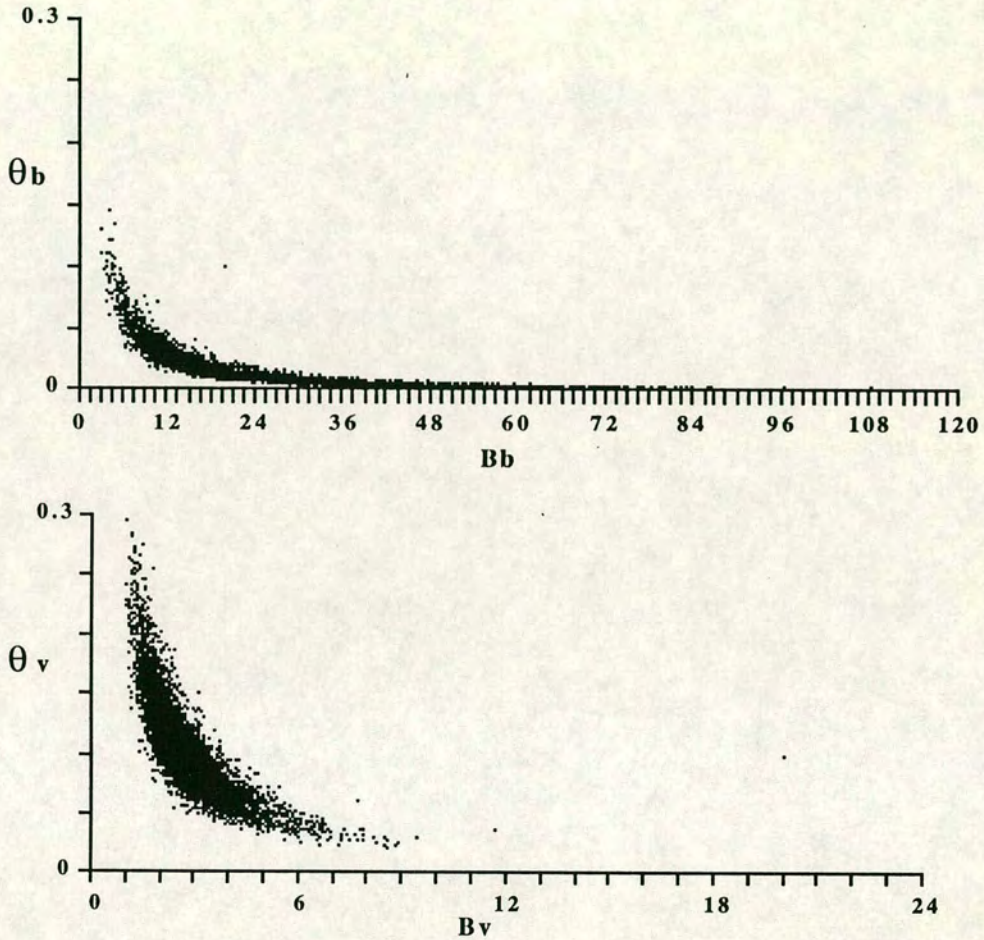


Fig. 3.4.8 Distributions of the likelihood of different cline shapes given a variance in gene frequency of F_{st}' 0.0068 (estimated from the discordance between loci). Each point represents one step in a random walk over 8000 replicates. The density of the distribution has a probability equal to the likelihood. The rate of decay (θ) and barrier strength (relative to width, i.e. B/w see text) are plotted on the *bombina* (top) and *variegata* (bottom) side of the zone. Note that the rate of decay can be strong and the barrier weak or vice versa.

The shape and course of the cline when the variance, F_{st}, is estimated alongside the other parameters

The most likely estimate of Fst

F_{st} was estimated as the variance of the marker loci from the average cline shape (F_{st}'; section 5, chapter 2). Instead of constraining F_{st} to a particular value it can be varied alongside the other parameters. This will provided an estimate of the variance over and above that expected from disequilibrium. In a similar manner to that above the Metropolis algorithm was set at T = 1 and the most likely estimate and limits were generated over 3563 replicate runs. This estimates F_{st} as 0.025 (limits: 0.015, 0.045, Table 3.4.3 Fig. 3.4.9). This implies that the mean gene frequency between sites fluctuates much more than that estimated from the discordance between loci where F_{st}' = 0.0068.

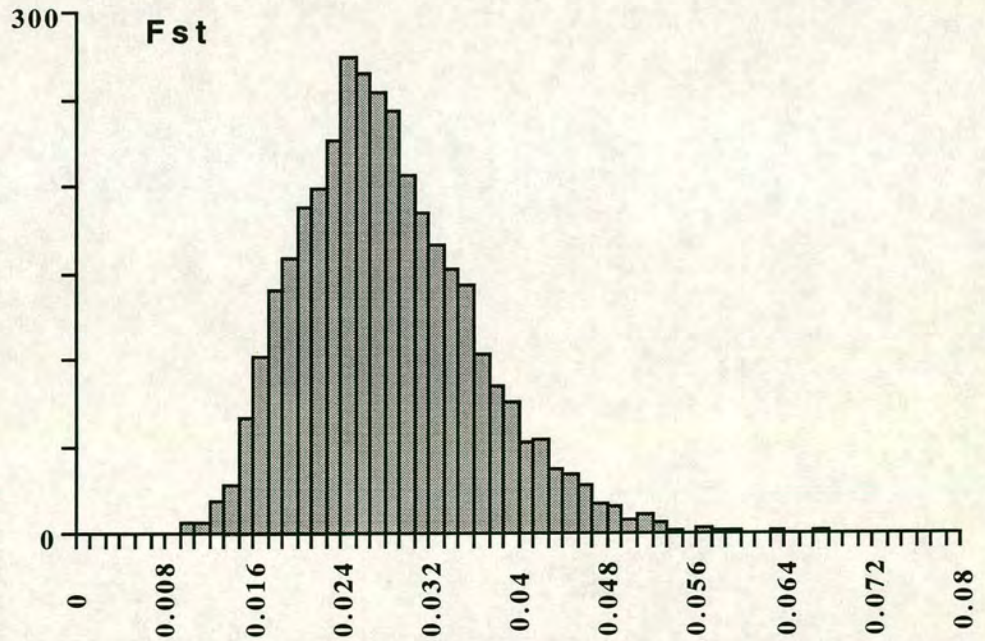


Fig. 3.4.9 The likelihood distribution of the variance in the mean gene frequency ($\bar{p}$) estimated as F_{st} between sites across the cline. The density of the distribution is proportional to the likelihood. The most likely value of F_{st} over 3563 replicate runs of the metropolis algorithm is 0.025. Limits on this estimate are bounded under the 2.5% area of the graph at either tail.

The shape and course of the cline when Fst = 0.025

The effect of this increased variance on the parameters describing the cline is to increase the limits around all the estimates (Table 3.4.3, Figs 3.4.10-12). The most likely estimates for the individual parameters also change. The rate of introgression on the *Bombina* side is no longer significantly different from 0 while the barrier strength on this side of the zone has increased to 40.85widths, but with much wider limits (6.83, 666.20). This means that the limits to the barrier strengths on either side of the zone now overlap and can no longer be held to be significantly different although the difference between the most likely estimates is actually greater. The limits and estimates of alpha have changed little indicating the robust nature of this estimate (α = 0.30 when $F_{st}' = 0.0068$ and 0.31 when $F_{st} = 0.025$). The standardised width has also changed little (1.079 compared to the previous estimate of 1.047) and the cline has only shifted slightly towards the *variegata* side of the zone (y = 0.015 compared to 0.006).

Estimating the likelihood of the one dimensional cline

The increased value of F_{st} means that each site effectively contributes less information to the position of the cline than when F_{st}' was set at 0.0068. A new effective sample size has to be estimated in the same way as Chapter two (eqs 2.8.1, 2.8.3). This allows the likelihood of the cline to be measured given that the likelihood for each site (i) is:-

$$\log L_i = Ne[\bar{p}_{obs}.\ln(\bar{p}_{obs}/\bar{p}_{exp})+\bar{q}_{obs}.(\ln \bar{q}_{obs}/\bar{q}_{exp})] \qquad (3.1)$$

where Ne is the effective sample size and $\bar{p}_{obs}$, $\bar{p}_{exp}$, $\bar{q}_{obs}$, $\bar{q}_{exp}$ are the observed and expected gene frequencies averaged over the diagnostic loci. This is essentially a G-test.

As this cline is described by fewer parameters than a cline in two dimensions, there are more residual degrees of freedom associated with it. The cline in one dimension is described by seven parameters (y, w, Bb, Bv, θb, θv, and α). There are 134 sites determining cline shape which means there are 126 degrees of freedom associated with it. Overall the likelihood of the cline when $F_{st}' = 0.0068$ is $\log L_{126} = -157.12$, and when F_{st} is 0.025 then $\log L_{126} = -118.49$ (Ne and individual site values are given in Appendix 3.3B). Therefore having allowed for more of the variance around the cline

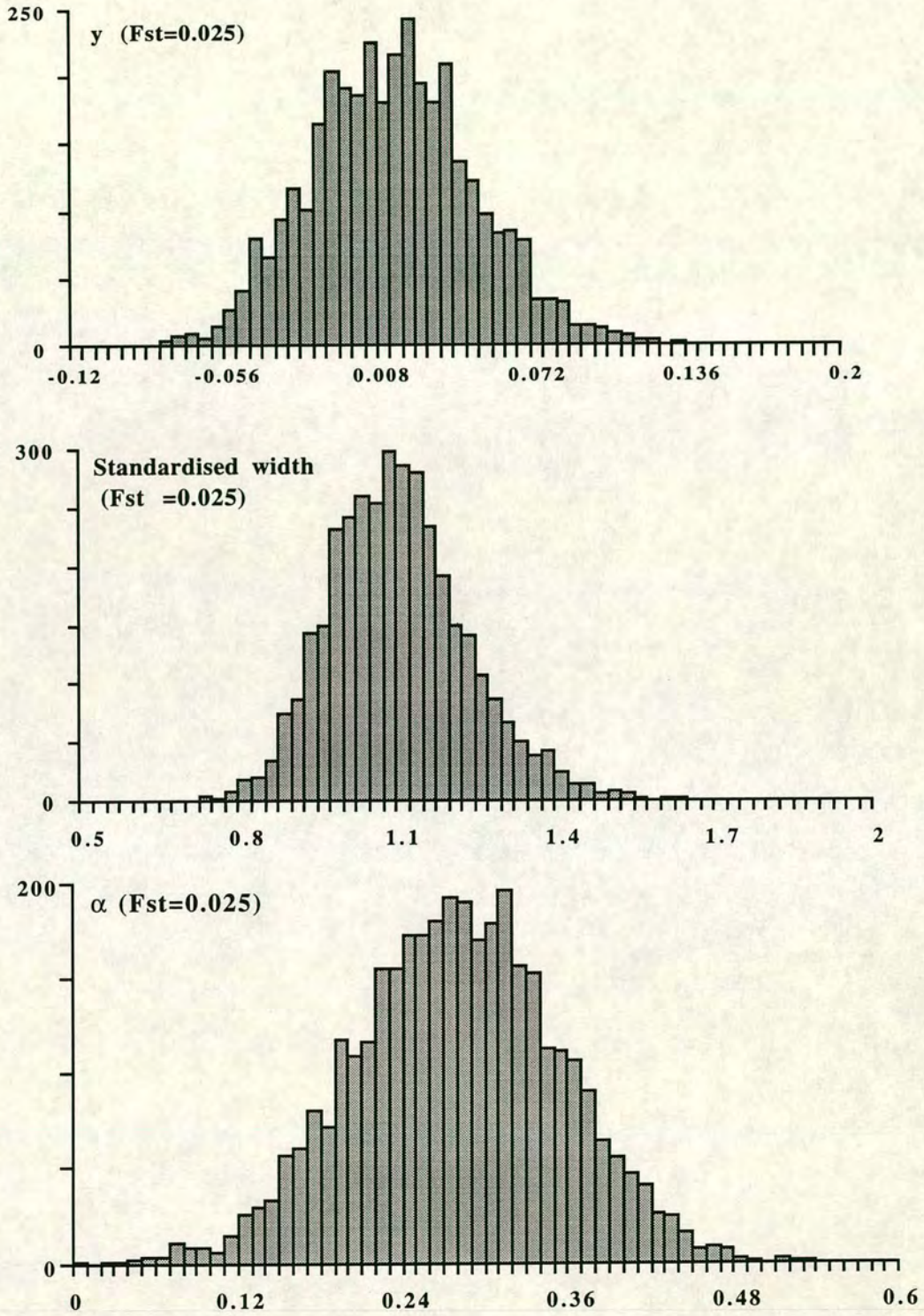


Fig. 3.4.10 The likelihood distribution of the position of the centre of the cline (y), the standardised width (w) and the difference in gene frequency by habitat (α). The density of the distribution is proportional to the likelihood. These estimates are made when the variance in gene frequency is estimated as $F_{st} = 0.025$. Limits are bounded under the 2.5% area of the graph at either tail.

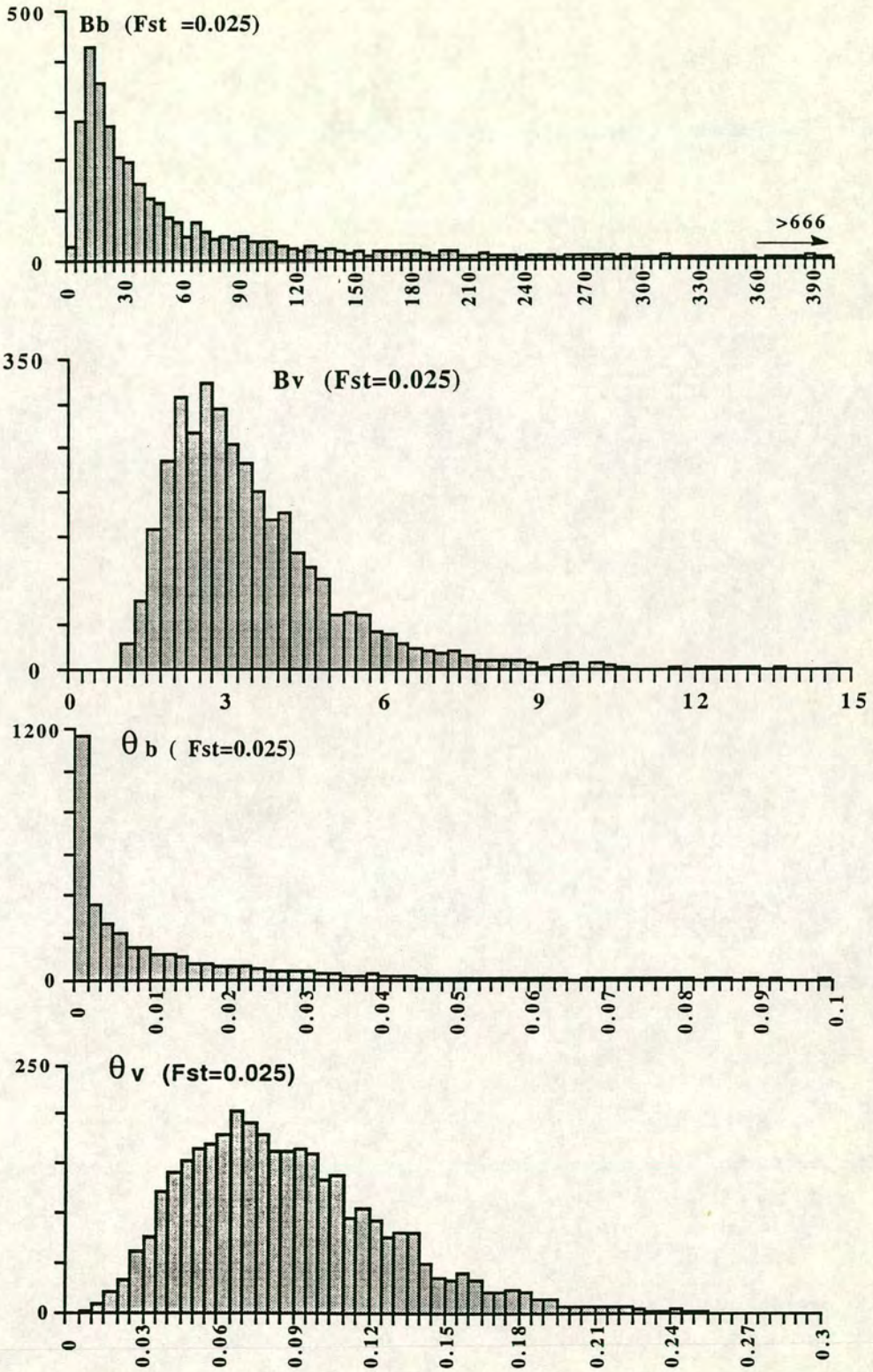


Fig. 3.4.11 Frequency distributions of the barriers to gene flow (B_b, B_v measured in widths, see text) and rates of introgression (θ_b, θ_v) either side of the zone when F_{st} is estimated as 0.025. The Metropolis algorithm was set at T = 1 for 3563 replicate runs creating a distribution whose density is equal to the likelihood. The limits bounded under the extreme 2.5% areas of the graphs are wider than when the variance is set at $F_{st}' = 0.0068$.

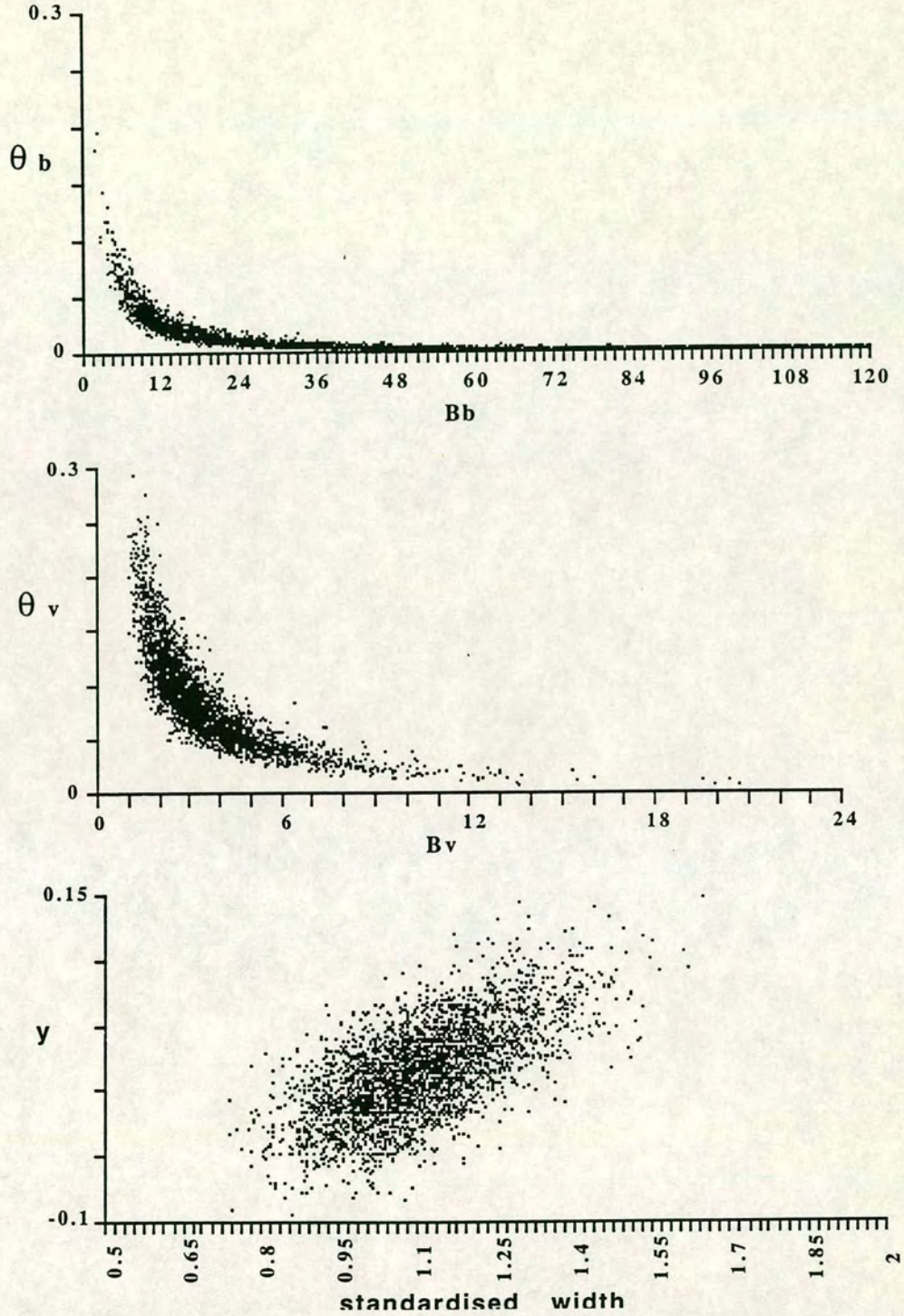


Fig. 3.4.12 Distributions of the likelihood of different cline shapes given a variance in gene frequency of F_{st} 0.025. Each point represents one step in a random walk over 3563 replicates. The density of the distribution has a probability equal to the likelihood. The rate of decay and barrier strength (measured in widths, see text) are plotted on the *bombina* (top) and *variegata* (middle) side of the zone. The bottom graph shows the position of the centre of the cline plotted against the width. The distributions are more scattered than when $F_{st}' = 0.0068$.

the likelihood has increased. However as $\chi^2_{126} = 152.92$ there is still residual variation even allowing for the increased $F_{st.}$ The model of the cline therefore does not fully account for all the variation between the sites. However most of the variation is accounted for by deviations at a few sites. There are twenty sites that deviate significantly from the expected gene frequency (i.e. whose likelihood is > 2, Fig. 3.4.13, Appendix 3.3). Of these only eight are significantly different at $p < 0.01$ (likelihood > 3.3).

The shape of the cline can be viewed by plotting the observed and expected frequency as a function of the standardised distance from the centre of the cline (Appendix 3.3B, Fig. 3.4.13). The gene frequencies are logit transformed ($p = \log[p/q]$). If the cline followed a smooth sigmoid curve then it would appear as a straight line.

The expected frequency of a population a certain distance from the cline differs between habitats such that $\bar{p} = \bar{p} + \alpha H \bar{p}\bar{q}$. This results in two clines; one through the puddle habitat (H = 1) and one through the pond habitat (H = -1). As $\alpha = 0.31$ the difference between the two clines in the centre is 0.15. The expected cline in ponds stops short as beyond this point (on the *variegata* side of the zone) there is no data to describe it. Both clines show a step in gene frequency ranging from ≈0.11-0.67. The shape of the clines in both habitats is the same. This was due to constraints of the program. As stated before the ideal situation would have been to estimate cline shape separately for each habitat type. This will be done at a future date.

Inferences from the shape of the cline will not be made in this chapter. Chapter 6 will discuss these in the light of the results of the following two chapters. However an explanation can be provided here as to of why α differs between that directly observed and that estimated using the metropolis algorithm.

The difference between the observed and estimated value of α.

Some of the residual variation around the cline might be accounted for if the difference between habitats is actually greater than that estimated from the model. The estimated value of $\alpha = 0.31$. This means that in the centre of the cline the difference in gene frequency between habitats is 0.15. The observed difference in gene frequency estimated directly is 0.5 (section 3.3). This is much larger than that estimated by the

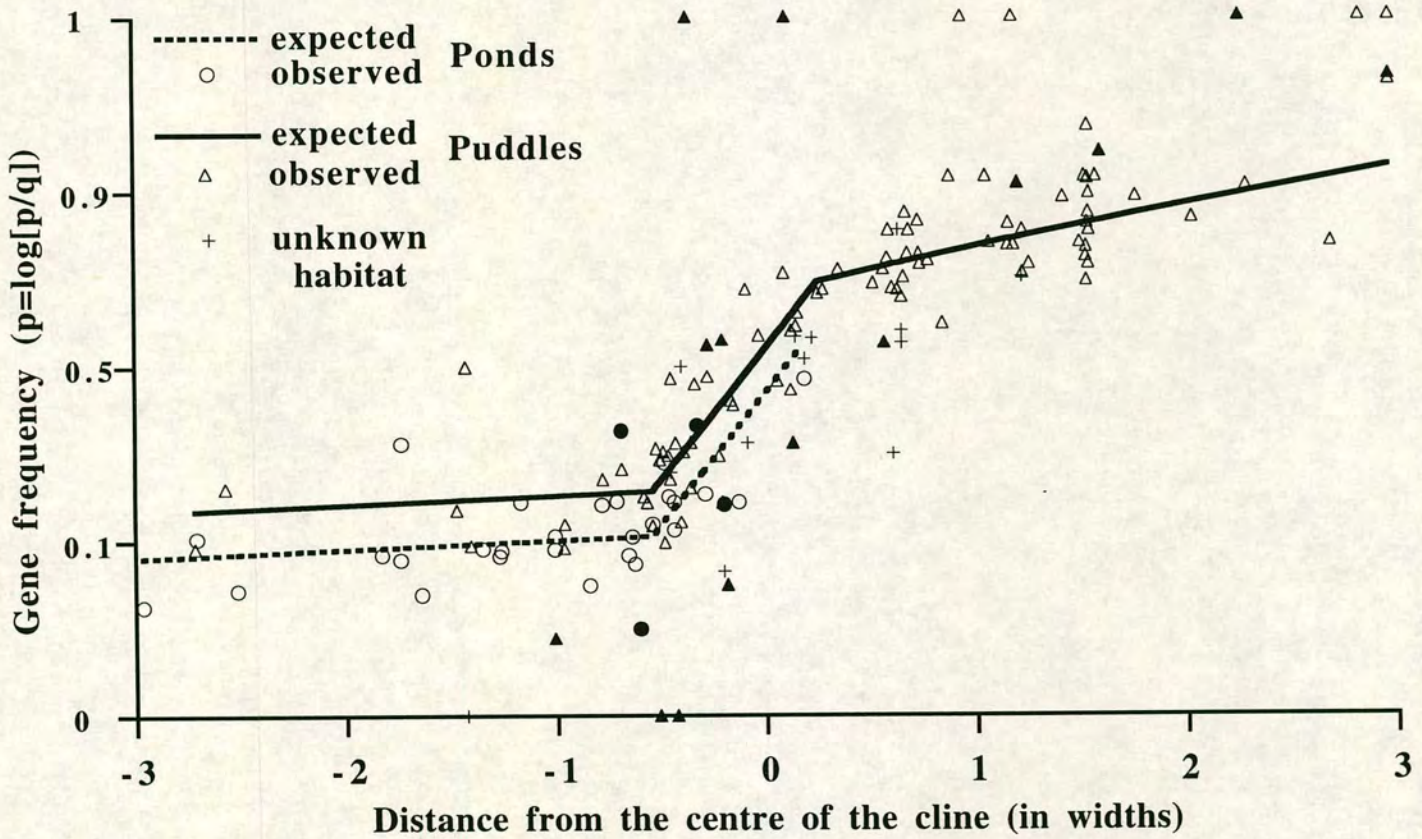


Fig. 3.4.13 The frequency of *variegata* enzyme alleles plotted on a logit scale (sites fixed at 0 or 1 are arbitrarily assigned to $\log_e[p/q] = \pm 5$). The expected frequency differs depending on the habitat type such that $\bar{p} = \bar{p} + \alpha H \bar{p}\bar{q}$ (where H = 1 for puddles and -1 for ponds). The difference in gene frequency between habitat types in the centre of the zone = 0.15 ($\alpha/2$). The observed gene frequencies are plotted for both habitat types and for those sites where the habitat is unknown (147 sites in total). The variance in gene frequency is estimated as $F_{s\,t} = 0.025$. The likelihood of this cline is $\log L_{126} = -118.49$ (see text for details).

model. The equivalent alpha would be 1! The discrepancy may be explained in the following ways:-

1. The model of alpha does not allow for any asymmetry in the difference in gene frequency between habitats across the cline. The barrier to *variegata* alleles may be stronger on the *bombina* side of the zone than *vice versa.* If this reflects 'extrinsic' selection then it is reasonable to assume that differences between habitats may be greater on the *bombina* side of the zone than on the *variegata* side.

2. It is likely that the model describing the cline incorporates more detailed geographic information about the cline than the simple assumption of looking at sites within a certain region of the study site. The regions used were large. If a comparison was made within a region that had many puddles and only a few ponds which were situated at the *bombina*-edge of the region, then the difference in gene frequency in that particular region would appear large. This would especially affect those regions near the centre of the cline.

3. The sample sizes of ponds used to observe the differences in gene frequency between regions is very small. It is likely there is a lot of variation around this estimate.

The true estimate of the difference in gene frequency between habitats in the centre of the zone is probably larger than that estimated by alpha but smaller than the observed estimate.

3.5 The pattern of genotypes, disequilibrium and heterozygote deficit in relation to the centre of the cline.

The analysis specifies the distance of each site from the centre of the cline (Appendix 3.3B). A summary of results from Chapters 2 and 3 is given in Table 2.6.3. These results are given for populations with five or more individuals. The sites are arranged in order of increasing distance from the centre of the cline i.e. from the *bombina* to the *variegata* side.

The distribution of the hybrid index clearly shows how the proportion of *variegata* alleles increases across the zone (Table 2.6.3). It also demonstrates the range of genotypes in populations near the centre of the cline. Most sites contain individuals of mainly one genotype. In the centre of the zone (≈X/W = -0.6 to 0.6 in table), sites contain predominantly one genotype but many have a few individuals of the opposite type. For example site 1055 is predominantly *bombina*-like yet two individuals have an average gene frequency of 0,88 and 1 each (7 and 8 *variegata* alleles respectively). This is reflected by the significant disequilibrium and heterozygote deficit at this site. Disequilibrium (measured as R) and the heterozygote deficit (F_{is}) increase in the centre of the cline (Fig. 3.5.1, values at each site are given in Table 2.6.3). This result confirms those in Chapter 2 where values increased for populations of intermediate gene frequency. Both disequilibrium and the heterozygote deficit are large in the centre of the zone. When there is a large scatter of points it is difficult to see any pattern. There is some suggestion that values of R and F_{is} increase more quickly on the *bombina* side of the zone.

Disequilibrium is significant for nineteen sites (Table 2.6.3). Eight of these sites also have a significant heterozygote deficit. These are all in the centre of the zone on the *bombina* side where X/W ranges from -0.5 to 0.05 (Table 2.6.3). Outside this distance only one other site shows an F_{is} significantly different from 0. This is also on the *bombina* side (site 1052). This is not just a reflection of sample size as there are many sites on the *variegata* side with large samples. Therefore this also implies that R and F_{is} are probably stronger on the *bombina* side of the zone.

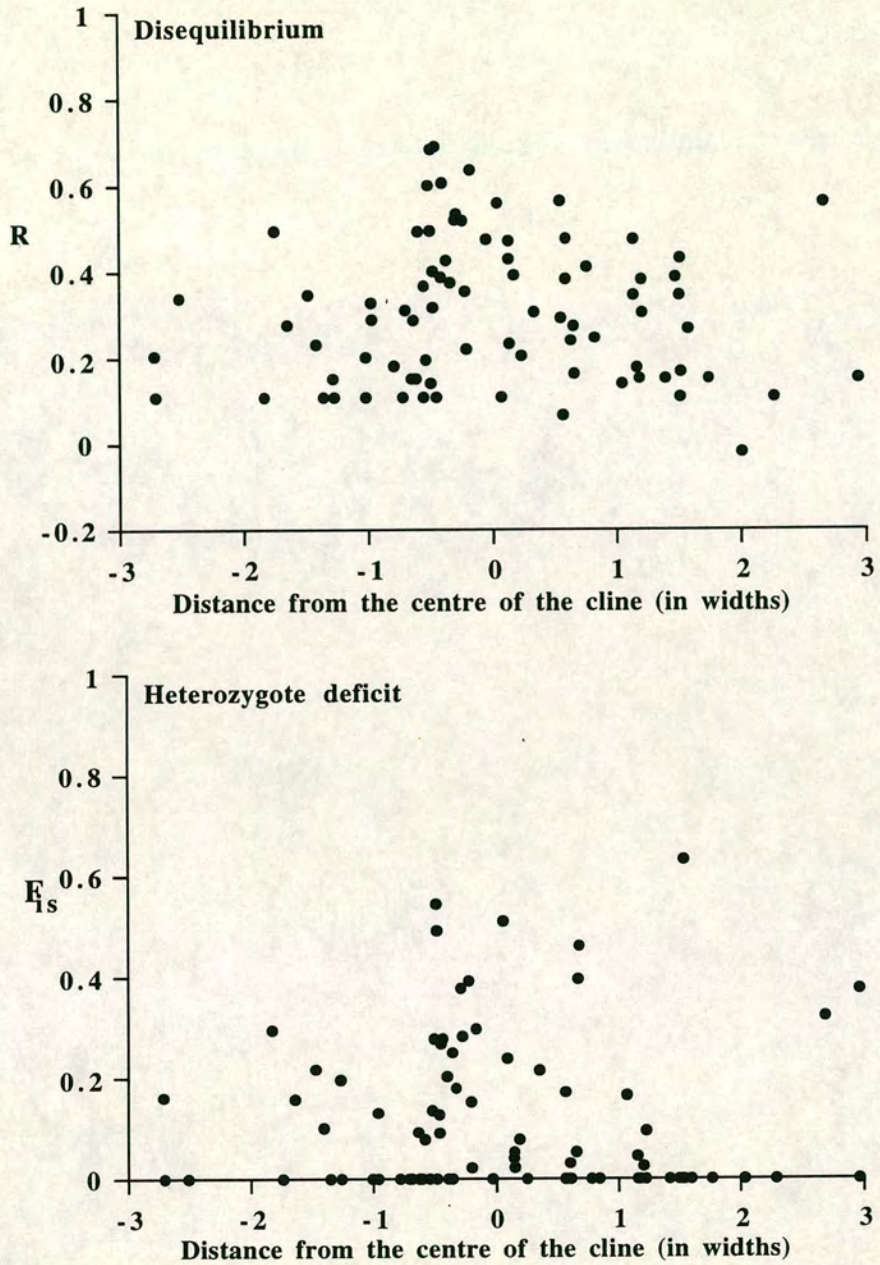


Fig. 3.5.1 The patterns of disequilibrium and heterozygote deficit across the zone. The distance of each site from the centre of the cline has been standardised by the width at that point. The cline allows for variable width and a difference in gene frequency between habitats (section 3.4). Only sites with five or more individuals have been included (N = 85). Values of the standardised linkage disequilibrium (R) and the heterozygote deficit (F_{is}) increase towards the centre of the zone.

Synopsis of results

1. The sites across the study area could be divided into different habitat types which showed a consistent correlation with the mean genotype of the population sampled from them. In general *bombina* like populations were found in 'ponds' while *variegata* like populations were found in 'puddles'. Populations tended to be sampled from ponds in the more arable regions of the study area while puddles were sampled from in forested areas.

2. The difference in gene frequency between the different habitat types was 0.5 when the average frequency across both habitat types was also 0.5.

3. The distribution of genotypes across the hybrid zone is best explained by a stepped cline where the width along the cline is variable and where there is a difference in gene frequency between habitats at the centre of the cline of 0.15.

4. There is some suggestion that the barrier to gene flow either side of the zone is asymmetric. The barrier to gene flow on the *bombina* side of the zone is greater than that on the *variegata* side.

5. The patterns of disequilibrium and heterozygote deficit in relation to the geographic centre of the cline are similar to results from Chapter 2. Both increase and reach a peak towards the centre of the cline. There is some suggestion that both are stronger on the *bombina* side of the zone.

Chapter 4

The relationship between dispersal, genotype and habitat

4.1 Introduction

Dispersal is an important parameter determining the distribution of genotypes in a hybrid zone (Chapter 1). For example, in dispersal dependent clines barriers to dispersal may stabilise the position of the hybrid zone (Hewitt, 1988). Where there is continual dispersal of parental genotypes into the zone then disequilibrium will be generated. These associations are continually broken down by recombination. When the recombination rate is known the strength of disequilibrium can be used to infer the dispersal rate (Chapter 1). Using this method the average dispersal distance for *Bombina* in Poland was estimated as 0.99km per generation at Przemyśl and 0.89km.gen$^{1/2}$ at Cracow (Szymura and Barton 1991, Table 1.2). This provides an objective measure of dispersal. Direct methods generally underestimate dispersal as the probability of observing long distance movement is small. For example the rate of dispersal from direct measurement at Cracow is estimated as 430m.gen$^{1/2}$ (Barton and Hewitt 1985, Szymura and Barton 1986).

The situation at Pešćenica is more complex than in Poland. Chapter 3 provided direct evidence for a close association between genotype and habitat. Other hybrid zones have demonstrated similar associations (Rand and Harrison 1989, Moore and Price 1993, Sites *et al.* 1994; Chapter 1). These all use selection to explain the relationship. At Pešćenica there are two explanations to account for this pattern.

1. Selection

There may be selection in relation to habitat such that adults not adapted are eliminated. Animals may disperse widely and at random in relation to the environment. Once they settle in a particular place there is then strong selection in relation to habitat. However it may be that selection is weaker and individuals actually disperse little. In this case associations between genotype and habitat will be built up over many years.

Habitat preference.

Individuals may not disperse at random. The association between genotype and habitat could be accounted for by an active habitat preference.

It may be that both selection and habitat preference contribute to the genotype/habitat association. The following chapter (Chapter 5) investigates whether genotypes are adapted to particular habitats. This chapter will consider the role of dispersal. An active habitat preference can only be demonstrated if evidence can be provided that individuals are choosing to move to different habitats according to their genotype. In order to demonstrate this individuals must have a range of habitat types within their dispersal distance and be able to move between them.

Recapture data can provide information about dispersal both in terms of distances moved and the nature of the individuals moving. The main aims of this chapter are to demonstrate how far individuals disperse and also to determine whether dispersal depends on genotype

4.2 Methods

Although individuals were toe-clipped for genetic analysis it was not necessary to use this as an individual mark. One of the more remarkable features of this animal is the highly variable colour of their belly. This aposematic coloration is displayed in the 'Unken' reflex, characteristic of *Bombina*. Their common names of fire and yellow-bellied toads reflect the difference between the two taxa. In general *Bombina variegata* has a large amount of interconnected yellow patches while *Bombina bombina* has smaller discrete red spots (Fig. 1.2). The number of connections between consistently placed spots in both taxa of *Bombina* has provided a convenient diagnostic score which is highly concordant with the enzyme markers when averaged across populations (Szymura and Barton 1986, 1991, Nürnberger *et al.* 1994.). Although the spotting pattern was recorded for every individual in this study an analysis of it will not be discussed here. More important to this study is the fact that the colour pattern is unique to every individual. This provides a reliable and non-invasive method of identifying individuals in the field.

It is surprisingly easy to catch individuals from sites such as puddles and shallow depressions. Individuals present can be identified some distance away as their eyes remain above the surface. Although their reactions are quick they do not respond as rapidly as many of the *Rana* species occurring at the same sites and are relatively easy to catch by hand. *Bombina* evade capture by diving into the soft mud at the bottom of the site (they never attempted to leave the site); however the disturbance of the substrate usually identified their position and they were caught more often than not. Even when they did hide they would often resurface within a short period of time providing another opportunity for their capture. I feel confident that at these sorts of sites we captured most of the population present in the water on any one day. Capturing individuals in deep ponds was much more difficult. Even when surrounded by a large chorus of animals it took some time to collect a sample. There are two disadvantages in a pond; *Bombina* seem to respond more quickly to water vibrations than they do to visual stimuli and when they dive there is little chance of retrieving them. Populations in ponds tend to be larger. This is reflected by the sample sizes of these sites. Recapture in one deep pond was attempted but the investment of time was great and with little reward; it was therefore abandoned. The recapture experiments were therefore carried out in puddles or depressions.

More than two thousand individuals were collected over the two field seasons in 1991 and 1992. Each individual was photographed, toe-clipped and released back into the site in which they were found either the same or subsequent day. Individuals were subsequently identified initially by the presence of a toe clip and then by their photograph.

4.3 Distances moved by individuals between sites.

Some sites were sampled repeatedly, either to increase the sample size of that population or to do detailed recapture studies. A number of individuals originally caught at one site were subsequently recaptured at another. Movements by individuals were picked up within the same field season and also between field seasons in different years (Table 4.3.1). In total, twentyseven individuals were recaught at sites different from their original (over the 1991 and 1992 field seasons). It is difficult to determine the pattern of movement by sex and/or genotype when sample sizes are

Site	Ind	sex	Moved to	Distance m (not less than)	$\bar{p}$ (mean gene frequency)	Time interval (days)
Recaptures within a field season (1991)						
1001	3	male	1003	1000	0.75	5
1001	18	female	1003	1000	0.62	9
1002	2	-	1001	1500	0.62	15
1002	8	-	1001	1500	0.88	39
1002	11	female	1003	150	-	12
1002	18	male	1003	150	0.67	19
1003	1	male	1001	1000	0.62	18
1054	1	male	1056	50	0.00	5
1054	6	female	1056	50	0.50	5
1054	8	male	1056	50	0.62	5
1054	9	male	1056	50	0.25	9
1054	14	female	1056	50	0.75	5
1054	16	female	1056	50	0.83	5
1054	18	male	1056	50	0.50	15
1054	22	male	1056	50	-	9
1054	23	female	1056	50	0.00	19
1054	25	male	1056	50	-	10
1054	30	female	1056	50	0.37	5
2003	8	male	2002	150	0.50	14
2003	17	male	2002	150	1.00	9
Average distance moved (standard deviation)			Total (n=20)	358 (516.1)		
			Males (n=11)	250 (373.5)		
			Females (n=7)	200 (354.7)		
			$\bar{p}$<0.5 (n=5)	50 (0)		
			$\bar{p}$»0.5 (n=12)	434 (553.09)		
Recaptures between seasons a. 1991/2						
1001	14	female	2002	1500	0.62	≈1 year
1001	17	male	2002	1500	1.00	≈1 year
1001	20	female	2002	1500	0.13	≈1 year
1002	5	-	2003	500	0.75	≈1 year
1002	15	male	2003	500	0.75	≈1 year
1003	7	female	2002	500	0.75	≈1 year
1045	1	male	2044	500	0.13	≈1 year
Average distance moved				929		

Table 4.3.1 Distances moved by individuals between sites. All the sites are of the same habitat type (2='puddle') The time interval reflects the time they were last seen at their original site and the first time they were seen at a new site. On average individuals move more over a year than within the three month field season. Males and females move a similar distance within a three month period. Individuals with a mean gene frequency, ($\bar{p}$), less than 0.5 move less than those with a higher gene frequency within the three months. However sample sizes are too small to say anything definite.

small. Also the probability of picking up individuals at different sites will depend on the frequency other sites are visited and the distance between those sites. However a number of conclusions can be made for the individuals that were observed to move:-

1. Individuals can move relatively large distances within a short time period; five days for individual 1001/3 to move 1km, 15 days for individual 1002/2 to move 1.5km.

2. The distance moved by individuals between years was larger on average than the distance moved within the field season. Within the three month field season of 1991 individuals moved 355m on average. The average distance covered in a year, from 1991 to 1992 was 929m.

3. Both sexes move between sites and cover a range of distances from 50m to 1.5 km (the latter detected between field seasons). On average females moved less than males but samples sizes are small and the standard deviation of each estimate is large.

4. Within the 1991 field season individuals with a mean gene frequency lower than 0.5, i.e. *bombina*-like individuals moved less far than those with a higher gene frequency. However again sample sizes are small; there were only four *bombina*-like individuals. Also the difference in movement is not reflected by individuals caught the subsequent year. Of the seven individuals caught the following year two were *bombina*-like and one of these had moved 1.5km. Therefore the range of distances observed are moved by individuals of all genotypes from *bombina*-like individuals to *variegata*-like ones.

It is known that individuals can move very long distances but it is rare that this movement is picked up. However one *bombina*-like individual caught in 1994 had moved 6.5km since it was last caught three years ago (L Kruuk pers. comm.). It moved from site 2012 on the *bombina* side of the hybrid zone to 1001 on the *variegata* side of the zone. This has important implications for measures such as disequilibrium as one or two *bombina* -like individuals in an otherwise *variegata*-like population can inflate estimates of disequilibrium at the edges of the zone (Szymura and Barton 1986, 1991; see also Chapter 1, and 6).

These results demonstrate that individuals of both sexes and involving a range of genotypes move relatively long distances. The toads are not large and considering the terrain they cover it is surprising they can move such distances in so short a time.

4.4 Mark recapture studies at particular sites

Methods

Detailed mark recapture studies were carried out at a number of sites in 1991 which varied in habitat and genotype (Table 4.4.1, Fig. 3.3.1). All the sites examined are reasonably close to the centre of the zone and are in a mixture of lowland forest and arable habitat (Chapter 3, Appendix 3.2, Table 2.6.3). Data were collected on a series of sampling days with a varying number of intervening days. New individuals that were caught were toe clipped and released either on the same day or the next day. They were available for recapture on the subsequent occasion. Therefore for the analysis the assumption will be made that they were released the same day. Individuals previously caught were identified in the field and released immediately. For one site (1001) the positions of all animals caught and recaptured within the site were noted. This allowed distances between the subsequent positions of animals to be estimated.

Site	$\bar{p}$	No of inds	No of sampling days	Starting date	Intervals between sampling (days)	Habitat type (discriminant score)	Distance from cline centre
1035	0.080	34	5	28/4/91	4,7,3,9	1 (3.03)	-0.64
1043	0.126	25	7	29/4/91	10,2,4,6,4	1 (3.93)	-0.71
1056	0.218	50	6	5/5/91	5,4,6,4,5	2 (-0.54)	-0.23
1044	0.236	24	6	29/4/91	10,1,4,6,4	2 (0.89)	-0.52
1064	0.252	36	4	10/5/91	4,10,6	2 (0.18)	-0.36
1054	0.594	34	7	1/5/91	4,5,4,6,4,5	2 (-1.08)	-0.21
1001	0.730	45	11	24/4/91	3,2,2,7,3,4,8, 7,2,4	2 (-1.08)	0.65

Table 4.4.1 Sites where recapture studies were undertaken. Recapture occurred over a number of sampling days with a varying number of days in between. The sites contain populations with a range of gene frequency, $\bar{p}$, and in different habitats. The different habitats are defined by the discriminant function (Chapter 3). The distance from the centre of the cline for each site is standardised by the width of the cline at that point (Chapter 3). No of inds is the number of individuals caught at a site.

Population sizes for each site was estimated using the Jolly-Seber stochastic method (Jolly, 1965; Seber, 1965; Seber, 1973). This model is the most biologically appropriate as it allows for variable survival rates. The only assumption it requires is that survival is age independent. Considering the longevity of these animals (≈ 10 years, Szymura and Barton 1986) and the brief sampling period (up to 42 days) this assumption is upheld. Recapture rates, population sizes, survival rate estimates (Ø) and estimates of the number of new animals arriving in the population (known as gains) can be made. Full details of the method are outlined elsewhere (Begon, 1979; Duellman and Trueb, 1986; Southwood, 1978). All the mark/recapture estimates were computed by the JOLLY program (Pollack *et al.* , 1990)

Given the short period over which the recapture experiments were done it is assumed that mortality is negligible. In this situation survival will reflect the rate of loss of individuals i.e. the emigration rate from a site. This is estimated as 1-Ø

Results

Recapture matrices

The results of the mark recapture studies can be displayed as a recapture matrix. A mark is made for every individual caught on a certain day and again if it is subsequently recaught another day. This means that one can easily trace the observed presence of individuals as they are caught and recaptured through time. Tables 4.4.2-8 show the results presented in this manner at all the sites. Similar patterns of individual movement are seen at all sites. There are some individuals consistently recaptured on subsequent sampling occasions, for example individuals 4, 6 and 21 at site 1001 (Table 4.4.2) or individuals 3, 20 and 24 at site 1054 (Table 4.4.3). Other individuals are caught once and never seen again (allowing for a number of sampling days since they were originally caught). At all sites apart from 1035 (Table 4.4.5), new individuals were continually caught on each sampling day. At 1054 the number of new individuals caught increases from two on the first sampling day to 15 on the last. If these sites contained closed populations or populations where there was little immigration then the numbers of new individuals caught over time would be expected to decrease. It could be argued that most of the population is not in the water but in surrounding crevices or under vegetation and as sampling occurred at different times of the day it is not surprising that there is a continual influx of individuals not seen

Individual	Sex	$\bar{p}$	Sampling day (date below) 1 21/4	2 24/4	3 26/4	4 28/4	5 5/5	6 8/5	7 12/5	8 20/5	9 27/5	10 29/5	11 2/6	Distance covered (m)	Distance gained (m)
1001/1	male	0.62	W				AO						*	166	166
1001/2	female	0.62	W			AN	W			F				350	38
1001/3	male	0.75	AX		W	AN	W	*						598	286
1001/4	male	1.00	AS			AN	W	CC	R	P	CC	*	*	340	6
1001/5	female	0.62	AS				W			*		P		304	304
1001/6	male	0.75	W				*	CC	FF	W				88	0
1001/7	male	0.88		W			AR2							242	242
1015/1	female	0.88		AT2											
1015/2	female	0.88		AT2			AR2		CC		*		*	283	283
1015/3	male	0.50		AT2		AN	AD						AR3	319	15
1015/4	female	1.00			AT2						AN	AR1		201	65
1015/5	male	1.00			AT2		AT2						AT1	1	1
1015/6	male	0.67			AT2					W	*	*		290	290
1015/7	male	0.83			AT2										
1001/8	male	-			W									176	176
1001/9	male	-			W										
1001/11	female	0.62			W						AP2	*	AD	230	122
1001/12	female	0.88			W	AN								156	156
1001/13	female	0.88			W	AN								156	156
1001/14	female	0.62			W										
1001/15	female	-			W				*					0	0
1001/16	female	0.68				AN				AP1				12	12
1001/17	male	1.00					R								
1001/18	female	0.62					W	*	*	*				0	0
1001/19	female	1.00					X	W			R	*	*	12	4
1001/20	female	0.12					X								
1001/21	female	-					W	CC	O	W			CC	62	6
1001/22	female	-					ZZ	VV						10	10
1001/23	female	-					AR2				AN	*	*	86	86
1001/24	male	-					AD			VV	UU	*		34	34
1001/25	female	0.75					X				AP2	AP1	AT1	294	294
1001/26	male	1.00					AY2	*	AY3	AX		AT2	AT1	8	8
1001/27	male	0.88					AT2			AX		AT2	AT1	3	1
1001/28	male	0.75					AT1								
1001/29	female	1.00					BF								
1001/30	male	0.75						AP4							
1001/32	male	0.88							BA						
1001/33	male	0.38							AN						
1001/34	female	-							BB						
1001/35	male	0.25							BB	W		*		10	10
1001/36	female	-								CC	W			14	14
1001/37	female	-								W					
1001/38	female	-								W		*		0	0
1001/39	male	-								AY2	*	AT2	*	7	7
1001/40	female	-									FF		P	34	34
1001/41	female	-									FF	*	AT1	274	274
1001/42	male	-									CC	*		0	0
1001/43	male	-									AR1	*		0	0
1001/44	female	0.88											AO		
1001/45	male	0.38											CC		
Number of new inds			6	4	12	7	23	9	11	16	15	17	17		

Table 4.4.2 Record of the date and position that individuals were caught at site 1001. The sex and mean *variegata* frequency is given for each individual when known. The name of the puddle that an individual was found in is denoted by the letters in the recapture matrix bounded by the double lines (Appendix 4.1 shows map and relative distances between puddles). * indicates that the individual has not moved sinces it was previously caught. The last two columns refer to recaptures only and are the total distance an individual covered and the distance it gained from its initial position within the sampling period. (N.B. the individual name is prefixed by the site number, the prefix 1015 is used as this site was initially distinguised from 1001 but was then incorporated into it.).

SITE 1054			Sampling day (date below)						
Individual	Sex	$\bar{p}$	**1** 1/5	**2** 5/5	**3** 10/5	**4** 14/5	**5** 20/5	**6** 24/5	**7** 29/5
1	male	0.00	*	*					
2	male	0.33	*			*			
3	female	0.83	*	*	*	*	*	*	*
4	male	0.50	*	*					
5	male	0.25	*						
6	female	0.50	*	*					
7	male	0.50	*		*				
8	male	0.62	*	*					
9	male	0.25	*						
10	male	1.00	*	*					
11	male	0.62	*						
12	female	0.62	*						
13	female	1.00	*						
14	female	0.75	*						
15	female	0.83	*	*					
16	female	0.83	*	*					
17	female	0.50	*						
18	male	0.50	*	*					
19	male	0.88	*						
20	female	1.00	*	*	*			*	*
21	female	-	*	*					
22	male	-	*						
23	female	0.00	*	*	*				
24	male	-	*	*	*	*	*	*	*
25	male	-	*			*			
26	female	-		*					
28	male	0.50		*	*	*	*	*	
29	male	0.50		*					
30	female	0.38		*					
31	male	0.50			*	*	*	*	
32	male	0.88			*	*	*	*	*
33	male	0.75				*	*	*	*
34	female	0.75						*	*
35	male	-							*
n			25	17	8	8	6	8	7

Table 4.4.3 Recapture data for site 1054. The sex and mean *variegata* gene frequency are given for each individual when known. A * in the recapture matrix (bounded by the double lines) indicates an individual was caught on that day. 'n' is the number of new individuals caught on any one day.

SITE 1056			Sampling day (date below)					
Individual	Sex	$\bar{p}$	1 5/5	2 10/5	3 14/5	4 20/5	5 24/5	6 29/5
1	male	-	*	*		*	*	*
1054/14	female	0.75	*	*				
2	male	-		*	*			
3	male	-		*	*			
4	female	-		*	*	*		
5	male	-		*	*	*	*	*
6	male	0.00		*	*			
7	male	0.75		*	*	*		
8	female	0.00		*	*	*	*	*
1054/1	male	0.00		*	*			*
1054/6	female	0.50		*	*	*		
1054/8	male	0.62		*		*		
1054/9	male	0.25		*	*		*	
1054/16	female	0.83		*	*	*	*	*
1054/22	male	-		*	*		*	
1054/30	female	0.38		*	*	*	*	*
9	female	0.12			*	*		*
10	male	0.00			*			
11	female	0.00			*			*
12	female	0.00			*			
13	male	0.17			*	*	*	
14	female	0.67			*		*	*
15	female	0.33			*	*		*
16	male	0.12				*	*	
17	male	0.00				*	*	*
18	male	0.38				*	*	*
19	male	0.00				*	*	*
1054/18	male	0.50				*		
20	female	0.00					*	*
21	male	0.50					*	*
22	female	0.50					*	
23	female	1.00					*	*
24	female	0.00					*	*
1054/17	male	0.50					*	
1054/25	male	-					*	*
25	female	-						*
26	female	-						*
27	male	-						*
28	male	-						*
29	female	-						*
30	male	-						*
31	male	-						*
32	male	-						*
33	female	-						*
34	female	-						*
35	female	-						*
36	male	-						*
37	female	-						*
38	male	-						*
1054/23	female	0.00						*
n			2	14	7	5	7	15

Table 4.4.4 Recapture data for site 1056. The sex and mean *variegata* gene frequency are given for each individual when known. A * in the recapture matrix (bounded by the double lines) indicates an individual was caught on that day. Individuals originally caught at site 1054 are given in bold. 'n' is the number of new individuals caught on any one day.

SITE 1035			Sampling day (date below)				
Individual	sex	$\bar{p}$	1 28/4	2 1/5	3 8/5	4 11/5	5 20/5
1	male	0.17	*	*			
2	female	0.17	*				
3	female	0.17	*	*			
4	male	0.00	*				
5	female	0.25	*	*	*		*
6	male	0.13	*	*			*
7	female	0.00	*	*			
8	female	0.00	*	*	*	*	
9	female	0.00	*				
10	female	0.00	*	*	*		
11	male	0.33	*	*			
12	female	0.13	*	*			
13	male	0.00	*		*		
14	male	0.13	*	*			
15	male	0.00	*	*			
16	male	0.13	*	*			
17	female	0.00	*	*			
18	female	0.13	*	*			
19	female	0.13	*	*			
20	male	0.00	*	*			
21	male	0.17	*	*			
22	male	0.00	*	*		*	
23	male	0.33	*	*			
24	male	0.00	*	*			
25	male	0.00	*	*			
26	male	0.13	*	*			
27	male	0.13	*	*			
28	female	0.00		*			
29	female	0.00		*			
30	male	0.00		*			
31	female	0.13		*			
32	female	0.13		*	*	*	*
33	male	0.00		*			
34	male	0.00		*			
n			27	7	0	0	0

Table 4.4.5 Recapture data for site 1035. The sex and mean *variegata* gene frequency are given for each individual. A * in the recapture matrix (bounded by the double lines) indicates an individual was caught on that day. 'n' is the number of new individuals caught on any one day; note that no new individuals are caught beyond the second sampling occasion.

SITE 1043			Sampling day (date below)						
Individual	sex	$\bar{p}$	**1** 29/4	**2** 8/5	**3** 9/5	**4** 10/5	**5** 14/5	**6** 20/5	**7** 24/5
1	female	0.17	*						
2	female	0.00	*				*		
3	female	0.00	*	*				*	
4	female	0.33	*						
5	female	0.25	*						
6	female	0.00	*						
7	male	0.13				*	*	*	
8	female	0.13				*	*		
9	male	0.00				*			
10	male	0.00				*	*		
11	male	0.13				*			
13	male	0.13					*		
14	male	0.13					*	*	
15	male	0.13					*		
16	female	0.13					*		
17	female	0.13					*	*	
18	female	0.25					*		
19	female	0.50					*		
20	male	0.00					*		
21	male	0.25					*	*	
22	male	0.00					*		
23	female	0.25					*		
24	male	0.00						*	
25	female	0.00							*
n			6	0	0	5	11	1	1

Table 4.4.6 Recapture data for site 1043. The sex and mean *variegata* gene frequency are given for each individual. A * in the recapture matrix (bounded by the double lines) indicates an individual was caught on that day. 'n' is the number of new individuals caught on any one day.

SITE 1044			Sampling day (date below)					
Individual	Sex		1 29/4	2 9/5	3 10/5	4 14/5	5 20/5	6 24/5
1	female	0.13	*					
2	male	0.37	*			*		
3	female	0.00	*	*			*	
4	female	0.25		*				
6	female	0.25		*				
7	male	0.50			*			
8	male	0.25			*			
9	female	0.00			*			
10	male	0.00			*			
11	female	0.37			*			
12	male	0.25			*			
13	female	-				*		
14	male	-				*		
15	female	-				*		
16	female	-				*		
17	female	0.25				*		
18	male	0.00				*		
19	male	0.50				*		
20	female	0.37				*		
21	female	0.17				*		
22	male	0.17				*		
23	female	0.00				*		
24	female	0.50						*
n			3	2	6	11	0	1

Table 4.4.7 Recapture data for site 1044. The sex and mean *variegata* gene frequency are given for each individual. A * in the recapture matrix (bounded by the double lines) indicates an individual was caught on that day. 'n' is the number of new individuals caught on any one day.

SITE 1064			Sampling day (date below)			
Individual	sex	$\bar{p}$	1 10/5	2 14/5	3 24/5	4 30/5
1	male	0.25	*	*		*
2	male	0.25	*			
3	male	0.88	*	*		*
4	male	0.37		*		
5	female	0.25		*		
6	female	0.00		*	*	
7	female	0.75		*		
8	female	0.33		*	*	
9	female	0.13		*		
10	male	0.37		*		*
11	male	0.13		*		
12	female	0.25		*		
13	male	0.00		*		*
14	female	0.88		*		
15	male	0.13		*		*
16	female	0.00		*	*	
17	male	-			*	*
18	male	-			*	*
19	male	-			*	*
20	-	0.37				*
21	-	-				*
22	-	0.13				*
23	-	0.50				*
24	-	0.25				*
25	-	0.88				*
26	-	0.00				*
27	-	0.37				*
28	-	0.00				*
29	-	0.13				*
30	-	-				*
31	-	0.00				*
32	-	0.13				*
33	-	0.13				*
34	-	0.00				*
35	-	-				*
36	-	-				*
n			3	13	3	17

Table 4.4.8 Recapture data for site 1064. The sex and mean *variegata* gene frequency are given for each individual. A * in the recapture matrix (bounded by the double lines) indicates an individual was caught on that day. 'n' is the number of new individuals caught on any one day.

before. However this would not account for the fact that some individuals were consistently recaught. The inference from this is that there are new individuals constantly moving into the site.

Conversely at site 1035 (Table 4.4.5) new individuals were caught on the first two sampling days only. The site was visited on three subsequent occasions at intervals of 7, 3 and 9 days respectively. Despite putting in a good effort to catch individuals very few were found. For example on sampling day four (11th May 1991) it took over 90 minutes to catch three individuals none of which was new. This implies that most of the population has left.

Jolly-Seber estimates

These results suggest that there is a continual turnover of individuals found at these sites. The Jolly-Seber estimates of population sizes, rate of loss and the gain of individuals for different sampling periods for each site reflect this (Table 4.4.9).

Sample sizes are small and the error is large for most of the estimates however there are still large differences between days at all sites. At site 1001, estimates of population size range from 27 individuals to 81 though the error for this latter estimate is very large. Site 1035 shows a decline in the population size whereas site 1056 shows an increase. Emigration and immigration rates fluctuate considerably within and between sites. For example there was 2% immigration at site 1035 between the first and second sampling days but between the second and third days 77% of the population emigrated.

Recapture rates also vary between sites. The overall probability of recapture was lowest for site 1001 (0.35) and highest for site 1054 (0.8). Averaging across all sites gives a total recapture probability of 0.61. It is surprising that the recapture rates (which allow for immigration and emigration) are so low considering the confidence in our ability to catch individuals. The proportion not caught may represent individuals at a site but not in the water.

Site	Sampling Interval	'Survival' Ø	Emi-gration 1-Ø	Population size	No of individuals gained	Recapture probability
1001	1	1.25 (0.38)	-0.25			
	2	1.05 (0.37)	-0.05	37.5	42.4	0.00
	3	0.81 (0.22)	0.19	81.6 (72.5)	-43.7 (57.9)	0.18 (0.08)
	4	0.78 (0.19)	0.22	22.1 (5.5)	18.4 (5.6)	0.31 (0.12)
	5	1.00 (0.21)	0.00	35.6 (6.3)	3.0 (6.6)	0.61 (0.13)
	6	0.89 (0.26)	0.11	32.7 (7.7)	12.9 (8.8)	0.27 (0.10)
	7	0.75 (0.19)	0.25	42.0 (12.1)	0.9 (7.0)	0.25 (0.10)
	8	0.71 (0.12)	0.29	32.4 (6.3)	4.6 (3.8)	0.48 (0.12)
	9	1.13 (0.15)	-0.13	27.5 (4.8)	3.2 (3.2)	0.53 (0.11)
	10			27.8 (5.3)		0.61 (0.13)
	Mean	0.93 (0.03)	0.07	37.7	3.7	0.35
1054	1	0.88 (0.24)	0.12			
	2	0.32 (0.12)	0.68	28.3 (7.9)	2.1 (1.6)	0.59 (0.18)
	3	0.78 (0.16)	0.22	11.0 (3.1)	0.8 (1.1)	0.70 (0.18)
	4	0.75 (0.15)	0.25	9.3 (2.7)	0.0 (0.34)	0.84 (0.14)
	5	1.00 (0.00)	0.00	7.0 (2.2)	0.0 (0.0)	0.86 (0.13)
	Mean	0.74 (0.06)	0.26	12.7	1.00 (0.4)	0.80
1056	1	1.00 (0.00)	0.00			
	2	0.98 (0.07)	0.02	16.0	7.8 (1.5)	1.00
	3	0.83 (0.12)	0.17	23.4 (2.0)	6.6 (2.8)	0.83 (0.11)
	4	0.77 (0.13)	0.23	26.2 (4.0)	7.7 (3.2)	0.63 (0.13)
	5			27.9 (4.3)		0.70 (0.12)
	Mean	0.90 (0.04)	0.1	23.4	7.37 (1.3)	0.79
1064	1	0.67 (0.27)	0.33			
	2	0.67 (0.39)	0.33	15.0	7.5 (6.8)	1.00
	3			17.0 (11.0)		0.30 (0.21)
	Mean	0.67 (0.24)	0.33	16.3	7.5	0.65
1044	1	1.67 (1.31)	-0.67			
	2	0.33 (0.26)	0.67	10.0 (10.0)	8.7 (7.1)	0.20 (0.24)
	3	1.62 (0.91)	-0.62	12.0 (6.8)	13.0 (19.6)	0.33 (0.27)
	4	0.11	0.89	32.6 (19.4)	0.4	0.54 (0.31)
	5			3.0		1.00
	Mean	0.93	0.07	14.4	7.1	0.52
1035	1	1.02 (0.17)	-0.02			
	2	0.23 (0.09)	0.77	35.4 (5.4)	0.1 (0.6)	0.84 (0.15)
	3	0.84 (0.57)	0.16	8.0 (8.0)	0.3 (0.2)	0.62 (0.21)
	4			7.0 (7.0)		0.43 (0.30)
	Mean	0.69 (0.19)	0.31	16.8 (2.9)	0.1 (0.6)	0.63 (0.2)
1043	1	0.33 (0.19)	0.67			
	2	1.50 (0.65)	-0.50	2.0 (0.7)	15.0	0.50 (0.35)
	3	0.80 (0.38)	0.20	18.0	6.9	0.00
	4	0.33	0.67	21.3 (8.6)	0.1	0.60 (0.28)
	5			7.0		1.00
	Mean	0.74	0.26	12.1	7.3	0.52

Table 4.4.9 Jolly-Seber estimates, over a series of sampling intervals, of survival rates, population size, number of new individuals gained and the probability of recapture for each site where recapture data were collected. Emigration is 1-Ø. When this value is negative it reflects immigration. Standard errors are given in brackets where possible. The total probability of recapture averaged across all sites is 0.61.

Comparing the observed and expected number of individuals caught

However sample sizes per day are small and the data are noisy. Immigration and emigration can be demonstrated more rigorously by plotting the proportion of new individuals expected to be caught against the observed proportion of new individuals actually caught on any one sampling day. This allows for differences in the total number of individuals caught but makes the assumption that each population is closed. The proportion of new individuals caught on any one day should decay with time at a rate determined by the recapture rate for that site. The expected value can be estimated given the overall recapture rate estimated for any one site using the Jolly-Seber method (Table 4.4.9).

For all of the 1991 sites apart from 1035 the observed proportion of new individuals caught on any one day was significantly different from that expected (Fig. 4.4.1). In general the proportion of new individuals caught on each day was greater. At 1035 the proportion caught is less than expected though this is not significant. This is reasonable if at 1035 the population emigrated after two days. Despite the high recapture rates for any one site new individuals are continually seen while some individuals are only seen once. At 1001 where the overall recapture rate is 0.35 the chances of not seeing an animal on any one day is 1-0.35=0.65. The chance therefore of not seeing an animal on eight consecutive occasions (e.g. individuals 1001/ 8 and 9) is 0.65^8 i.e. $\approx$ 1 in 31 . The conclusion must be that these animals have left.

Although this is strong evidence that the population movement at these sites is dynamic, more explicit evidence comes from the fact that animals not recaught at one site were subsequently caught elsewhere. At site 1056, twelve individuals were caught that were originally found nearby at site 1054. Site 1054 is 50m from 1056. This represents 34% of the population caught at 1054. Although the site is close by, it is still a large fraction of the population. Of these 12 individuals 6 were seen at 1054 on the previous sampling date. This suggests first that an individual not seen at this site may have left rather than hidden and second that they can move quickly to another site. Seven of the individuals arrived together on the same day (10/5/91).

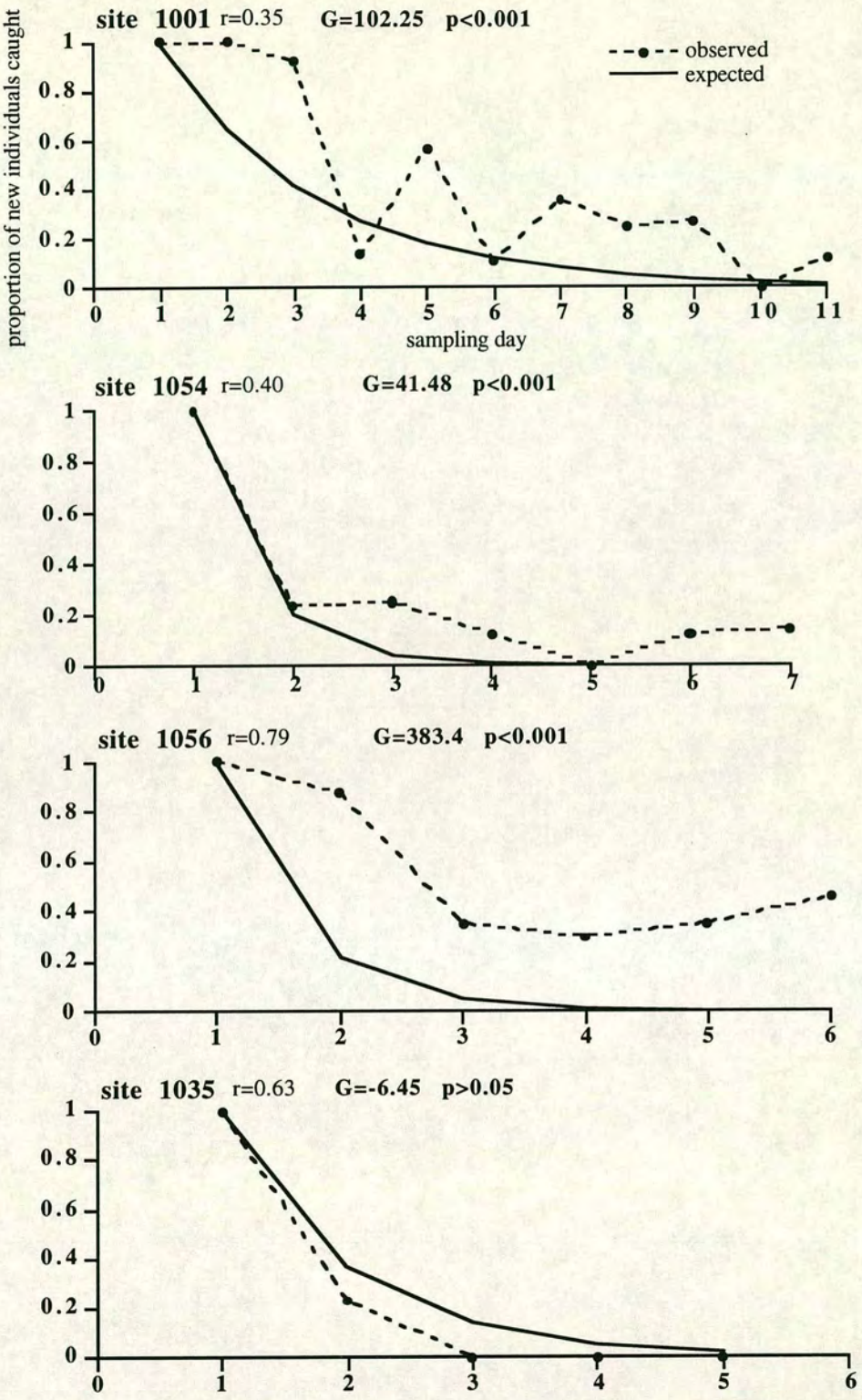


Fig. 4.4.1 and (overleaf) The proportion of new individuals expected to be caught on any one sampling day and that actually observed for sites sampled in 1991. 'r' is the overall recapture probability estimated for each site using the Jolly-Seber Stochastic method. G test statistics are given for each graph (degrees of freedom = no. of sampling days -2). The observed values are significantly different from the expected at all sites except 1035.

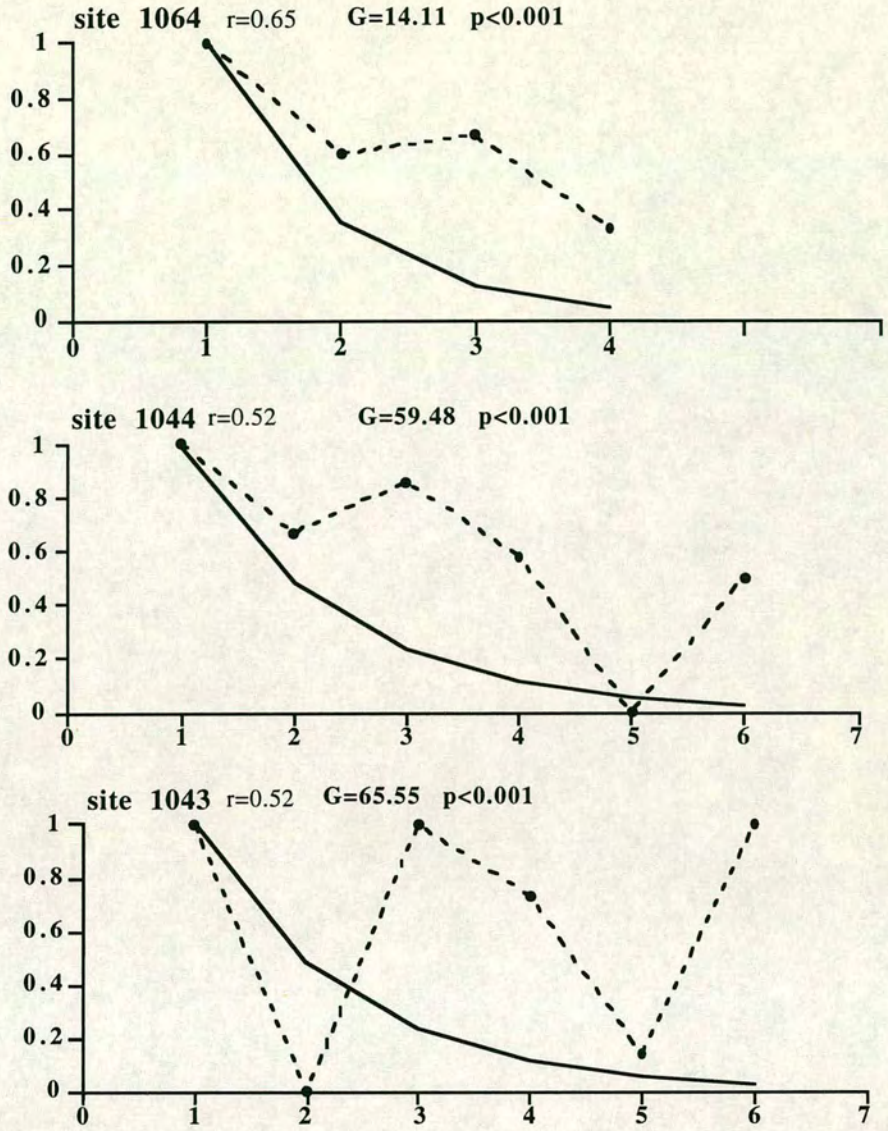


Fig. 4.4.1 continued

The relationship between habitat, genotype and dispersal

It is known that there is a correlation between the mean gene frequency of a population and the habitat in which it is found (Chapter 3). If there is a continual turnover of individuals through sites then is the relationship between habitat and genotype maintained over time? If so, this would imply that individuals were actively choosing habitats. Some sites, like those discussed here, were returned to a number of times and the mean gene frequency is therefore for a population caught at different times. However the majority of sites were sampled once. If the gene frequency of a population varies with time then sampling on only one occasion may not reflect the true nature of the population which comes into that site.

Sites 1056 and 1054 are 50m apart. Although they are both classified as habitat type 2 their discriminant scores differ. The score for 1056 is higher than that for 1054 being -0.54 and -1.08 respectively indicating it is more 'pond' like than 1054. The overall gene frequency for each site is $\bar{p}$ = 0.218 for site 1056 and 0.594 for site 1054 i.e. site 1056 which is more pond-like contains the population which is more *bombina*-like. The range of gene frequencies at each site reflects this (Fig. 4.4.2) Considering the turnover of animals through these sites and their proximity it is surprising that there is such a large difference in gene frequency. It is known that 12 animals from 1054 moved to 1056. The gene frequency of ten of these animals (the remaining two were not scored) ranges from 0.00 to 0.83. Overall the mean of this moving sample is 0.43, less than the overall population mean for 1054. It is tempting to suggest that the more *bombina* like individuals are moving from the puddle to the more pond-like site. However some of these individuals are very *variegata* like and 0.43 is still larger than the mean of the 1056 population at $\bar{p}$=0.218. Recapture was done over the same time period at each site except that the first sampling day at 1056 was the second at 1054. If the gene frequency of the population on each equivalent sampling date is compared at each site a consistent difference in gene frequency between these sites is maintained (Fig. 4.4.3). Therefore despite movement between these sites and presumably immigration from unknown sources as well, the relationship between the sites is upheld. On all sampling days the mean gene frequency of the population at 1056 is less than at 1054 and their error bars do not overlap except on sampling day 3.

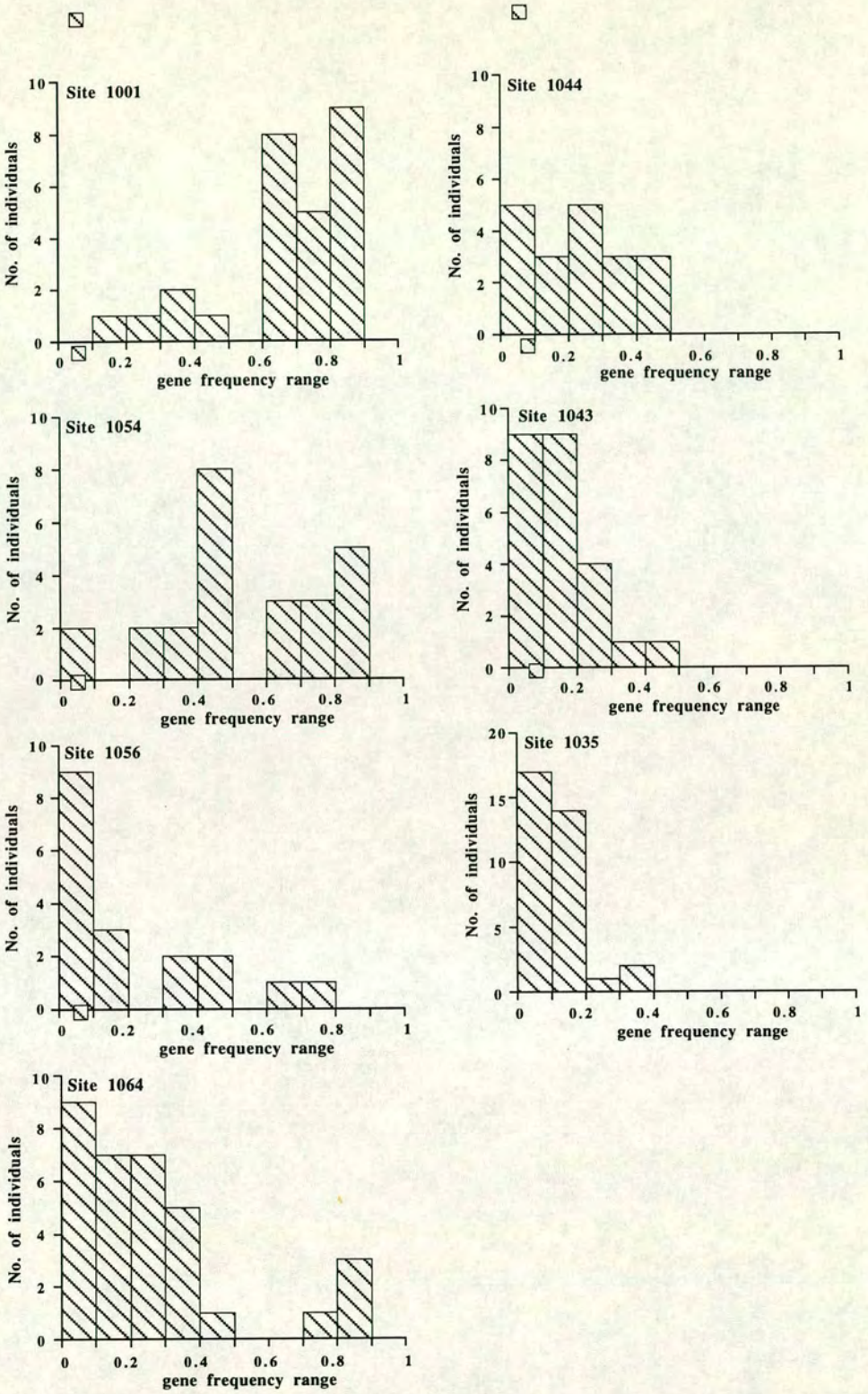


Fig. 4.4.2 Histograms showing the range of $\bar{p}$ across individuals at sites where recapture was undertaken.

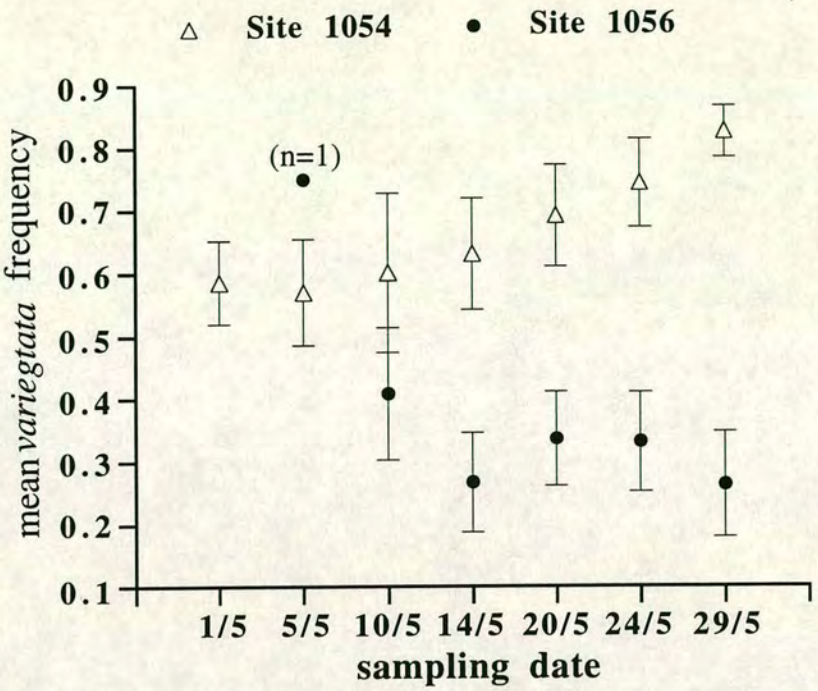


Fig. 4.4.3 The mean gene frequency (and standard error) of the population caught on different sampling days at sites 1054 and 1056. The two sites are 50m apart. The discriminant score of 1056 implies that it is more pond-like. The gene frequency at this site is consistently lower than at 1054 therefore it contains the more *bombina*-like population. This relationship is maintained over time despite a continual turnover of individuals and the movement of individuals from 1054 into 1056 (mainly on 10/5).

The majority of individuals from site 1054 were caught on 10th May at site 1056. The gene frequency of the population is elevated on that day compared with the following three. There is no evidence that the more *variegata*-like individuals of the moving sample then left after that date. Indeed individual 1054/16 ($\bar{p}$=0.83) stayed as long as individual 1054/1 ($\bar{p}$=0.00) yet overall the population at 1056 has the lower gene frequency. Therefore despite some variance this result suggests that although both sites are available to all individuals the more *bombina*-like individuals are consistently found in the more pond like site.

This is a comparison of only two sites. The strength of the comparison relies on the fact that these sites contain populations with a high variance in gene frequency, are close to each other and there is proven movement between them. However other sites show the same consistency in gene frequency across sampling days (Fig. 4.4.4).

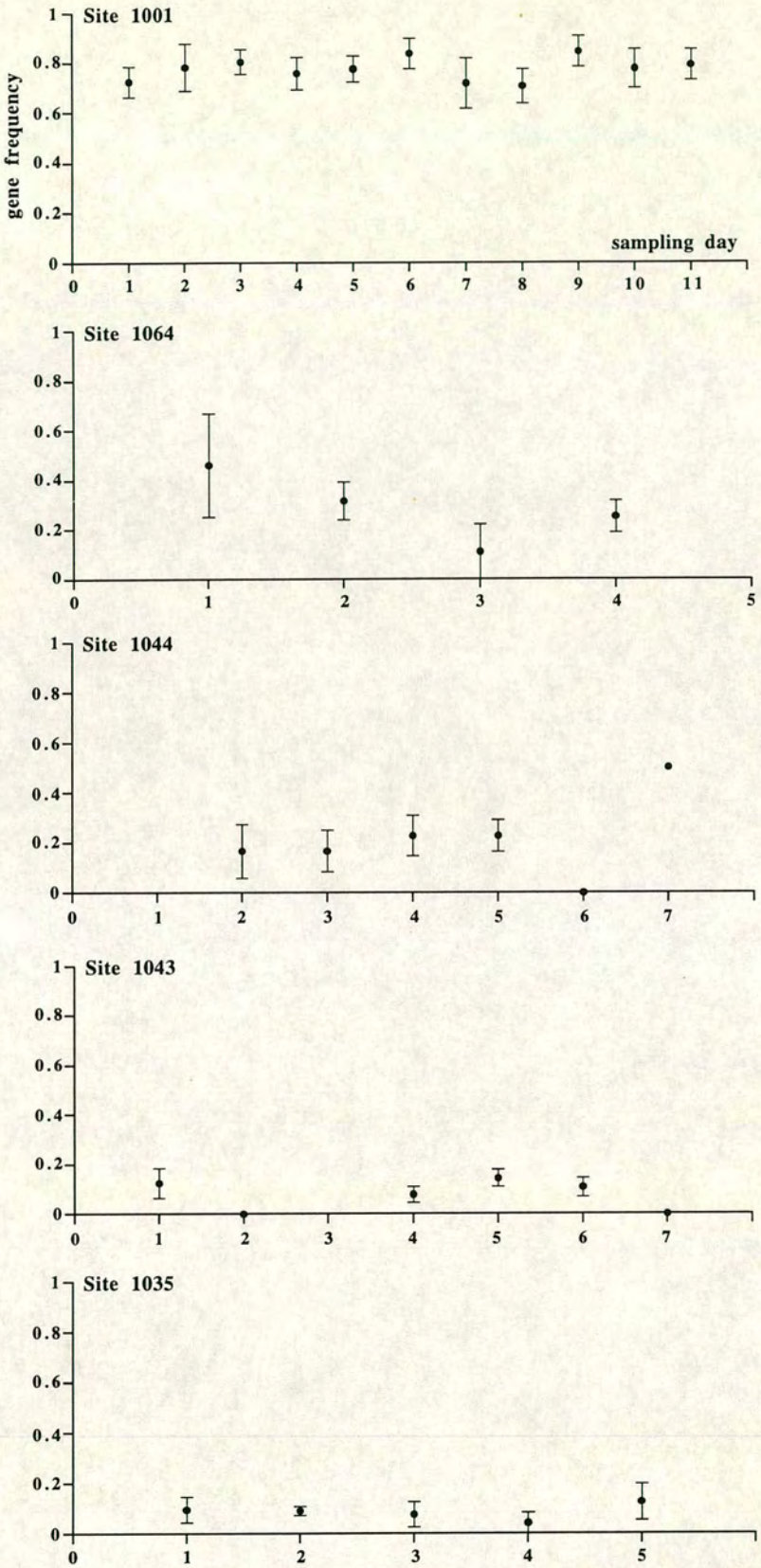


Fig. 4.4.4 The mean *variegata* gene frequency of the population caught on any one sampling day at five sites. Within each site the gene frequency remains relatively constant over time.

However this result has to be treated with some caution because the variance in gene frequency at these sites tends to be less (Fig. 4.4.2).

It could be argued that the reason that the gene frequency at site 1056 is consistently less than 1054 is because there is a large influx of *bombina*-like individuals from a source that is closer to 1056 than 1054. Indeed there is a pond, site 1055, which is 200m from 1056 and 250m from 1054; i.e. 1056 lies between the two (Fig. 3.3.1). It has a discriminant score of 3.95, which is at the extreme 'pond' end of the function, and $\bar{p}$=0.139. Twenty three animals were caught here (recapture was not attempted at this site). Not one of these animals was caught at either 1056 or 1054 throughout the month when recapture was undertaken. Although the population size is not known for this site it seems unlikely that no individuals would be recaught if they were moving between sites. Furthermore beyond this site, again at a distance of approximately 200m lies a more *variegata*-like site, 1064 where $\bar{p}$=0.252. Recapture was done at this site (Table 4.4.8) but again no individuals from the pond were picked up. This result has two implications; it suggests either that individuals from a pond habitat do not emigrate to puddle habitats or that there is very little movement out of pond sites containing relatively pure *bombina* individuals. Both explanations may be involved although it is known that *bombina*-like individuals do disperse relatively long distances (Table 4.3.1). Differential dispersal will be discussed below.

Although 1056 and 1054 differ in their discriminant score, suggesting that one is more pond-like than the other, they are both of the same habitat type. All the recaptures observed between sites are between the same habitat types (Table 4.3.1). This may provide evidence that individuals are faithful to a particular habitat type but is more likely to reflect the fact that puddles were sampled more often than ponds. However different habitats are available to individuals. The distances between 1056, 1054, 1055 and 1064 are within the range of distances that different genotypes were observed to move (Table 4.3.1).

The evidence so far suggests that individuals are actively selecting particular habitat types. However this sort of analysis highlights the problem of disentangling habitat use and habitat availability. The immediate discussion above involves four sites of different habitat types and containing populations with a different mean genotype. Although the area was searched diligently for other sites none in the immediate region were found. Nevertheless there is the possibility that some sites were missed and the consistency of gene frequency across sampling days may be proportional to random

immigration of genotypes from other unidentified sites. Even so, I think these results present strong evidence for a habitat preference. What must be remembered is that all this analysis is based on the relationship between habitat and the distribution of enzyme markers. As these genes are not presumably responsible for the preference itself, the preference genes must be in linkage disequilibrium with them. Therefore, the preference itself must be stronger than revealed here. This will be discussed further in Chapter 6.

Differential dispersal in relation to sex and genotype.

The most detailed recapture data are for site 1001 (Table 4.4.1). The movements of 50 individuals were recorded over 42 days on eleven sampling occasions. The site consists of a network of puddles on a track bordering the edge of the forest and a field. The site extends for 400m along the track. On each sampling day every puddle was checked for toads and the position of each individual was recorded. This site contains a very *variegata* like population, overall $\bar{p}=0.73$. For those individuals whose genotype have been scored only five out of 27 have a mean gene frequency lower than 0.6, the remainder are between 0.62 and 0.88 (Fig. 4.4.2). Such low variation means that it is difficult to see if there is any pattern of movement by genotype within this site. However it provides an opportunity to record the movement of *variegata*-like animals within a restricted area.

The distance moved by any individual can be measured in two ways, either as the total distance travelled or by the distance gained from where it started. As individuals are first caught on different sampling days the distance is standardised by the number of days in which the movement occurs (although this makes the assumption they move in a linear direction). Table 4.4.2 gives both distance measures for each individual. The distances moved by individuals are highly variable (Fig. 4.4.5). Considering that these distances have been standardised by the number of days over which the movement occurred the distances covered are large. The median distance moved by individuals is 5.67m per day and the distance gained is 1.97m per day (Table 4.4.10). This of course does not reflect the dispersal out of the site. It seems adaptive for animals to move around a site such as this, basically a large collection of puddles. The movement was recorded during the breeding season and the presence of eggs and tadpoles were noted. Considering the ephemeral nature of puddles it is reasonable to presume that

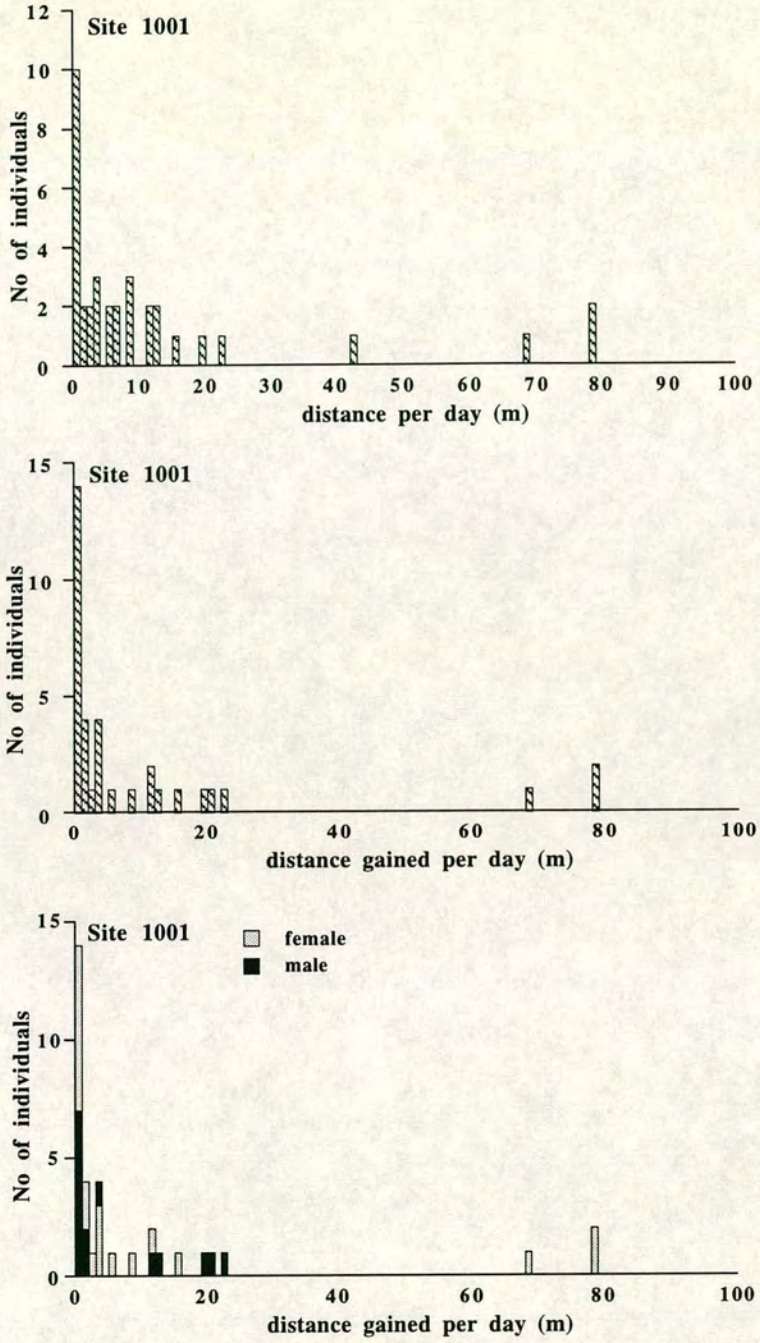


Fig. 4.4.5 The distances moved by individuals at Site 1001. The top graph shows the total distance moved by individuals around the site. The middle graph shows the distance each individual has gained from the position it was originally caught from. The bottom graph shows the distance gained by males and females in the site. All distance measures are standardised by the number of days over which the movement occurred.

sex	No of individuals	Median distance moved per day (m)	Median distance gained per day (m)
Female	19	5.67 (0, 78)	3.29 (0, 78)
Male	16	4.68 (0, 42)	0.82 (0, 22)

Table 4.4.10 Total distance moved and distance gained per day by males and females at site 1001. The minimum and maximum observed values are given in brackets.

individuals move around in order to maximise their reproductive success. There are no differences between the median distance moved per day between males and females however the median distance gained by females in a day is greater than that for males. This difference however is not significant (Mann-Whitney U=125.5; p=0.38).

The difference is not significant but there is some suggestion that although males and females cover the same distance the females may travel further i.e. males move within a more restricted area. Evidence from other studies of *Bombina* suggest that males of *Bombina bombina* are territorial (Aguade *et al.* , 1992). Although *Bombina variegata* are cited as not being territorial, it may provide an explanation for the difference in movement; males may stay in one area more than females. It also makes sense for females to move further and lay eggs in different puddles.

Certainly at this site individuals move around a great deal. This is a very *variegata*-like population despite a few *bombina*-like individuals. If *Bombina bombina* are more territorial and occupy more permanent habitat types it may well be that they not only move around less within a site but also between sites. With this limited data set it is extremely difficult to demonstrate this. However a comparison can be made at sites where there is a relatively large spread of genotypes, between the length of time individuals stay at a site and their genotype.

Comparison of the time different genotypes remain in a site.

At sites 1056 and 1054 the mean gene frequencies of the populations are consistent over time and yet there is a difference in gene frequency between the two sites and a large spread of genotypes within the sites. The individuals sampled in each site can be divided into those with a mean gene frequency of $\bar{p}$ »0.5 or those with a genotype < 0.5 i.e. more *variegata*-like or more-*bombina* like. If information is combined across both sites then the cumulative rate of decay of *bombina* and *variegata* individuals from these sites can be estimated. If there is no difference between the length of time that *variegata* or *bombina*-like individuals stay then there should be no difference in the decay rate. Estimates are made over the first six sampling days only as site 1056 was visited only six times. The assumption will be made that if an individual is not seen on a certain day but is subsequently seen, then it was present on the day it was not seen. The time interval between sampling days alternates between four and five days however because sample sizes are small, the decay rate will be estimated as a function of sampling day rather than real time. There is an obvious problem that individuals caught on the last sampling day in each site will only have the opportunity to be present on one sampling day. However as this will apply to both genotypes it should not bias the result.

The results show that *variegata*-like individuals leave the sites at a faster rate than *bombina*-like individuals (Table 4.4.11, Fig. 4.4.6). This result must be treated with caution due to the small sample size and the small difference in the slopes. Also the length of time an individual stays may depend on how long it has been at the site prior to sampling. If, for example, this analysis was done on the population at 1035 it would be observed that most of the population would have left after two sampling days despite it being a very *bombina*-like population. Also, the sex of the individual has not been taken into account here. This may further increase the variance if, for example, *bombina*-like females move at the same rate as *variegata* like males. I think all that can be reasonably concluded from this analysis is the possibility that dispersal rates may not be the same either between the sexes or in relationship to genotype.

		Sampling days stayed (no of individuals)						
site	Genotype	1	2	3	4	5	6	Total No. of individuals
1054	»0.5	7	8	1	2	2	0	20
	<0.5	3	1	1	1	0	0	6
1056	»0.5	3	3	3	1	1	0	11
	<0.5	3	4	4	4	3	0	18
Cumulative Total	»0.5	31	22	10	6	3	0	31
	<0.5	24	18	13	8	3	0	24
decay rate	v-like	1	0.70	0.32	0.19	0.10	0	0
	b-like	1	0.75	0.54	0.33	0.13	0	0

Table 4.4.11

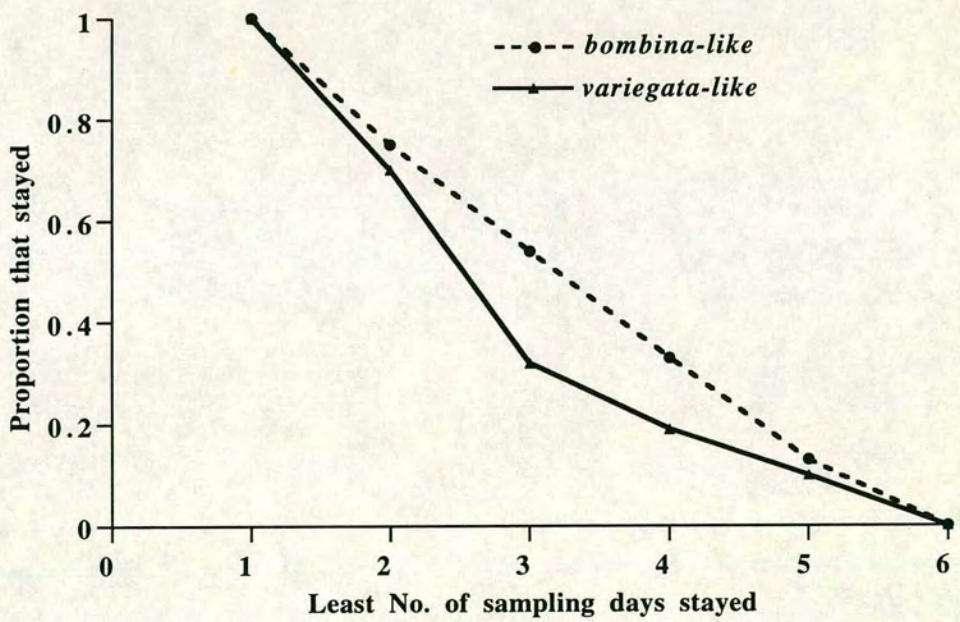


Fig. 4.4.6

Table 4.4.11; Fig. 4.4.6 The rates of decay of *bombina* and *variegata*-like individuals from sites 1056 and 1054. Individuals with $\bar{p}$ »0.5 are *variegata*-like; those <0.5 are *bombina*-like. The cumulative rate of decay of each type is plotted in Fig. 4.4.7. These results suggest that *variegata* may disperse more quickly than *bombina*.

Synopsis of Results

1. Individuals of all genotypes and both sexes disperse. Movements observed ranged from 50m to 1.5km within a field season.

2. The average distance moved by individuals within the three months of the field season was 358m while the distance moved between years averaged 929m

3. There was an observed turnover of individuals at some sites.

4. Despite some observed movement through sites the gene frequency of each site and the correlation between habitat and frequency remained relatively consistent between sampling days.

5. Sites of different habitat type were available to individuals of different genotype. The distances between these types were well within the range of movements observed. This and the consistency of gene frequency at these sites suggest that individuals of particular genotypes chose particular habitat types. Therefore *bombina*-like individuals tended to prefer ponds whereas *variegata*-like individuals prefered puddles.

6. There is some suggestion that females moved around more than males within sites and that *variegata* dispersed at a greater rate than *bombina* though these results have to be treated with caution.

The most important result to emerge from this chapter is the evidence of a habitat preference. This has important implications for direct estimates of disequilibrium, indirect estimates of dispersal rates and selection and the nature of the barrier to gene flow. These issues will be addressed in Chapter 6.

Chapter 5

The adaptive significance of a habitat preference

5.1 Introduction

Chapter 3 demonstrated a correlation between the mean genotype of the population and the habitat it is situated in. In general, populations in puddles have a higher *variegata* gene frequency than those in ponds. Chapter four provided evidence for dispersal between sites. Despite a turnover of individuals through some sites the mean genotype of the population remained similar between sampling days and the relationship between genotype and habitat held even though there was a high variance of genotypes in the area. This implied that adult toads were actively choosing different habitats according to genotype.

The aim of this chapter is to find whether the habitat preference is adaptive? There are two explanations for how a preference might arise:-

1. A habitat preference may exist as a result of conditioning. If two populations or taxa are associated with different habitats then they may tend to select the natal type or one they are used to as they do not recognise the other. Evidence for the effect of experience on habitat preference usually comes from comparison of wild caught populations with manipulated reared ones for example the effect of laboratory rearing on the habitat preference in deermice (Wecker, 1963).

2. A preference may arise if it confers a fitness advantage on the individuals expressing it. There are many models to demonstrate this, mostly investigating the possibility of sympatric speciation (Diehl and Bush, 1989; Jaenike, 1988; Rausher, 1984; Rice, 1987; Wilson and Turelli, 1986). It is often empirically difficult to demonstrate that a habitat preference is adaptive although there are some cases (Partridge, 1978; Pulliam and Danielson, 1991).

Often active habitat selection may be confused with passive habitat sorting, for example variation in morph frequencies in the cricket frog *Acris crepitans* is more likely due to differential mortality between habitats than a preference (Nevo, 1973).

Whether the preference is learned or adaptive is important for understanding the nature of selection happening within this hybrid zone; i.e. whether it is against hybrids or is due to adaptation to different environments (Chapter 1).

Laboratory studies indicate that *variegata* lay fewer, larger eggs that develop more quickly than *bombina* (Nürnberger *et al.* , 1994; Rafinska, 1991). This is a well known phenomenon of amphibian populations at higher altitude compared to more lowland ones, (Berven, 1981) and reflects the upland habitat of *variegata*. Larger eggs are generally a result of colder temperatures while the decreased development time may be an adaptation to puddle habitat given their ephemeral nature (Seidel, 1987). These results indicate that *bombina* and *variegata* are adapted to different environments. However it is difficult to extrapolate from the laboratory to the natural environment. Field experiments are needed for a more comprehensive assesment of differential selection in puddles and ponds where there is competition between the opposing taxa. There are a number of questions that can be addressed in this context:-

1. What is the relative fitness of pure *bombina* and pure *variegata* in each habitat.

2. What is the fitness of hybrids with different proportions of *variegata* and *bombina* alleles in each habitat.

3. What is the fitness of eggs from a hybrid population in different habitats.

All these three fitness measures may vary. Practically it would be a daunting task to measure the fitness of a range of genotypes in different habitats across the hybrid zone. The following experiment focusses on the relative fitness of pure *bombina* and *variegata* tadpoles in puddles and ponds on either side of the zone. Relatively pure *bombina* and *variegata* eggs are translocated to a different habitat type on the opposing side of the hybrid zone and reared together in equal quantities. Measurements of survival, growth and development time will be assessed for each genotype in relation to habitat.

5.2 Methods and materials.

Egg collections

The experiment was conducted in June 1992. The experiment coincided with the observed onset of breeding. Suitable sites for the translocation experiment were monitored before its onset. Although eggs had already been observed in ponds none were initially seen in puddles. The experiment began once eggs were laid in puddles. Large numbers of eggs were laid in the puddles over a relatively short period. All the eggs required for the experiment were collected between the 24th and the 26th of June.

Eggs were collected from a number of sites on either side of the hybrid zone which were known to contain relatively pure adult populations (Table 5.2.1). As these sites are far from the centre of the zone it is assumed that the genotype of the eggs will reflect that of the adult population. *B. bombina*-like eggs were collected from two large ponds (1039 and 2116 Fig. 3.3.1) where the mean frequency of *variegata* alleles, $\bar{p}$ = 0.062 and 0.081. They are both classified as type 1 habitats and have discriminant scores of 4.82 and 3.04 respectively. They are therefore at the extreme pond-like end of the discriminant function (Chapter 3). *Bombina variegata*-like eggs were collected from a series of puddle sites in upland forest near Perkoveć (Fig. 3.3.1). The mean gene frequency of these populations range from 0.813 to 0.968 and their discriminant scores range from -1.29 to -1.55. When each set of sites is combined the distribution of the gene frequencies are well separated by a central gap (Fig. 5.2.1). Therefore the two sets of sites differ in both genotype and habitat. Two of the puddle sites (2131 and 2137) have no ecological data associated with them and the adult population in the immediate area was not genotyped. However they are in close proximity to the other puddle sites and as these sites do not vary much in either habitat or genotype it is assumed that they are similar.

Eggs had to be collected from many more *variegata* than *bombina* sites as puddles are much smaller than ponds and fewer eggs are found in them. Female *bombina* lay eggs in a number of clusters. Before collecting these eggs the sites were initially scanned to get an overview of the distribution of egg clusters. Whole clusters of eggs were collected from different regions of the sites. Not all clusters from a particular area within a site were collected. It was hoped that allowing the sampling to be as

Site	Distance from cline centre	Habitat type discriminant score	$\bar{p}$	No. of egg clusters	Total no. of eggs collected	Mean no. of eggs in batch (sd)	Total survived till hatching
Bombina bombina sites							
2039	-1.27	1 (4.82)	0.062	45	849	18.9 (10.1)	835
2116	-1.01	1 (3.84)	0.081	52	1185	22.8 (11.1)	1184
Total				97	2034	21.0	2019
Bombina variegata sites							
2122	1.57	2(-1.39)	0.938	9	120	13.3 (12.5)	119
2124	1.53	2(-1.46)	0.813	2	32	16.0 (17.0)	31
2126	1.52	2(-1.29)	0.828	6	41	6.8 (3.9)	41
2127	1.53	2(-1.41)	0.968	10	256	25.6 (14.0)	255
2131		2		16	220	13.8 (9.5)	220
2132	1.52	2(-1.45)	0.938	4	28	7.0 (4.2)	28
2133	1.54	2(-1.55)	0.922	4	33	8.3 (5.6)	33
2134	1.53	2(-1.45)	0.886	12	148	12.3 (14.2)	147
2136	1.54	2(-1.55)	0.900	32	558	17.4 (12.9)	556
2137		2		6	69	11.5 (5.2)	68
2138	1.54	2(-1.41)	0.875	11	130	11.8 (6.2)	130
Total				112	1635	14.6	1628

Table 5.2.1 Sites where eggs were collected for the translocation experiment. Eggs were collected from two large ponds (habitat type 1) and eleven puddle sites (habitat type 2). The discriminant score of each site is given in brackets after the habitat type (Chapter 3). The mean *variegata* gene frequencies, $\bar{p}$, show that the pond sites contain virtually pure *bombina*-like individuals whereas the puddle sites contain almost pure *variegata* individuals. The distance from the centre of the cline (standardised by the width) is given for each site. The number of egg batches and total number of eggs collected and surviving to hatchling are given for each site and overall (see text).

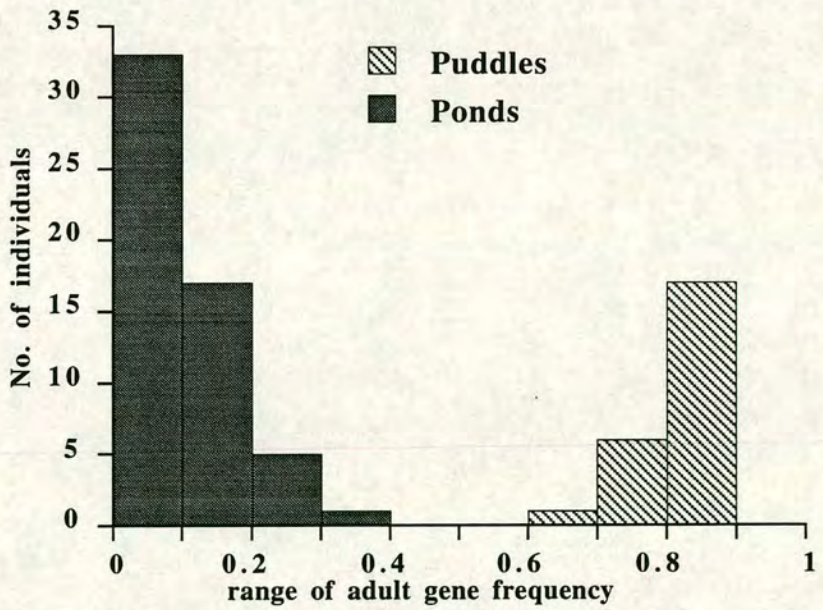


Fig. 5.2.1 The range of gene frequencies of the adult populations combined across each set of sites.

widespread as possible would minimise the chance of collecting a large number of eggs from one family. Overall 2034 eggs were collected from the pond sites while 1635 eggs were collected from the *variegata* sites. These were distributed over 97 and 112 egg clusters respectively; in general the clusters of eggs from the *bombina*-like sites contained a larger number of eggs (Table 5.2.1).

The diameter of five eggs were measured from a subsample of the batches at most sites using a dissecting microscope fitted with a graticule (Appendix 5.1, Table 5.2.2). Measurements were estimated to the nearest 0.02mm. The average diameter of each batch was converted to an estimate of egg volume under the assumption that eggs are spherical. The mean egg volume was estimated for each site and combined over sites of similar habitat (Table 5.2.2). Overall the volume of *variegata* eggs is 4.5 times bigger than *bombina* eggs.

Site	Number of egg batches	Mean Egg Volume mm^3 (standard deviation)
Pond sites		
2039	19	1.83 (0.20)
2116	15	1.54 (0.38)
combined	34	1.71 (0.32)
Puddle sites		
2122	11	7.38 (1.06)
2126	4	6.73 (1.49)
2127	8	9.29 (2.04)
2131	16	7.33 (0.74)
2132	4	8.91 (2.06)
2134	12	7.71 (1.64)
combined	55	7.79 (1.55)

Table 5.2.2 Mean egg volume from a number of batches from sites where eggs were collected from the translocation experiment. The diameters of five eggs from each batch were measured. The mean diameter of each five was converted to volume and mean volume averaged over all batches.

Egg clusters were reared separately to hatchling stage in small plastic picnic cups containing ≈250mls of tap water. This kept the families separate and ensured that any eggs that were not fertilised or that did not develop for some other reason were not included in the experiment. However most eggs survived to hatchling (Table 5.2.1).

Assigning individuals to enclosures

Once the eggs had reached the hatchling stage, they were distributed across 22 larger containers which would then be transferred to the field enclosures. Each container contained a mixture of 72 eggs collected from the *bombina*-like pond sites and 72 from the *variegata*-like puddle sites. Therefore, 1584 eggs from each of the two sets of sites were used in the experiment. In order to minimise family effects, clusters of eggs were split into batches of three. Each batch of three was assigned to a different container until all the eggs from one cluster were used up. As some clusters were larger than others it meant that more containers had representatives from that family than others.

The enclosures

The aim of this experiment was to release a mixture of tadpoles from the two taxa into both habitat types and compare the development and survival of both in each. There were 22 batches of mixed eggs. Of these, ten were assigned to ponds and twelve to puddles (Table 5.2.2) The two types of sites, puddles and ponds, present different sorts of problems in this sort of experiment. It would be ideal if the mixtures of tadpoles could be released back into puddles and ponds without interfering with the habitat at all. However, given the size of the ponds and the amount of flora and fauna found there we would never be able to retrieve the individuals at the end of the experiment. Different strategies had to be adopted.

Site	Region	Altitude	No. of enclosures within site (name)	No. of max-min thermometers	Co-ords x, y
PONDS					
Veleševeć pond	Veleševeć	98	4 (V1-V4)	2	≈8.32, 0.69
1039	Veleševeć	98	3 (V5-V7)	1	8.32, 0.69
Vratovo	Bulge	99	3 (V8-V10)	1	≈4.07, -4.28
PUDDLES					
2124	Perkoveć	209	1	0	-6.51, -5.58
2126	Perkoveć	209	4 (P1-P4)	1	-6.06, -5.43
2133	Perkoveć	209	1	1	-5.22, -5.34
2134	Perkoveć	209	1	1	-5.62, -5.39
2139	Perkoveć	209	1 (P6)	0	-5.80, -5.40
2140	Perkoveć	209	3	1	-5.81, -5.39
2141	Perkoveć	209	1	0	-5.73, -5.37

Table 5.2.2 The distribution of enclosures at different pond and puddle sites either side of the hybrid zone. The altitude and region of each site are given. Co-ordinates are measured from the arbitrarily assigned origin described in Chapter 2. Max-min thermometers were placed at most sites. The names of the enclosures where results were obtained are given in brackets (see Results section for details).

Pond enclosures

The only way to ensure a return of individuals from the ponds was to physically enclose them. Ten bags were made from fine nylon mesh. Each bag was 2 metres high and one metre in diameter. They were supported by chicken wire held in place by branches of hazel (Fig. 5.2.2). Stones were placed in the bottom of the bags to prevent them floating to the surface.

The ten bags were distributed between three ponds (Table 5.2.2). Ideally they should have been placed back into the sites where the eggs were collected. However site 2116 dried up to such an extent in a spell of hot weather (after the egg collections) that there was not enough water to support the enclosure bags. Bags were placed in site 1039 and in two additional more permanent sites. One was a large pond in the centre of Veleševeć village and the other was a large vegetated drainage canal in a cleared area of forest to the north of Pešćenica, called Vratovo. Although we never caught adult *Bombina* from either of these sites we did hear a chorus of toads from the Veleševeć pond and *bombina* individuals had been caught by Szymura in Vratovo in 1979 (site 6 Appendix 3.2).

Fig 5.2.2 Example of three pond enclosures in Vratovo.

Puddle enclosures

The remaining twelve containers were emptied directly into puddles formed by wheel ruts (Fig. 5.2.3). The individuals put into puddles did not have to be enclosed in bags as the puddles were small enough to be drained at the end of the experiment. The main problem with puddles was to prevent other toads from the area entering and laying eggs within them. In order to prevent contamination each experimental puddle was first checked for the presence of tadpoles and cleared of all existing eggs. The timing of the experiment was such that no eggs had been laid prior to the experiment. The puddles had previously been monitored as potential sites and eggs were collected as soon as they were laid. The puddles were then checked every 3-4 days for any newly laid eggs. As each take approximately a week to develop this ensured that new eggs could be removed before they hatched. Rainfall caused further problems at the puddle sites due to the potential of flooding. These sites were in upland forest which were subject to extreme and localised thunder storms. The weather was closely monitored and if a thunder storm was imminent the banks of the puddles were built up.

The experimental puddles were in the same area as the puddle egg collections. They were distributed across a number of sites. Four of the puddles were in site 2126, three at site 2140 and one each at 2124, 2133, 2134, 2139 and 2141. Again, the positioning of puddles had to be chosen with care. 2124, 2126, 2133 and 2134 were on well used tracks in the forest. They were ideal sites for the translocation experiment as there were many toads at each and eggs had been collected from them. However because of their situation they were at risk of being disturbed by tractors. Therefore the remaining sites, 2139, 2140 and 2141 were wheel ruts on a disused railway line (essentially a grassy bank) within the forest. *Bombina variegata* adults and eggs were found at these sites but not in the same quantity as the others.

The mixed batches of hatchlings were placed in the enclosure puddles on the 3rd of June 1992 and in the pond enclosures on 5th. The puddles were monitored at intervals of 3 and 4 days alternately. The pond enclosures were checked on the 15th and 23rd of the month. The maximum and minimum temperatures were noted on each occasion.

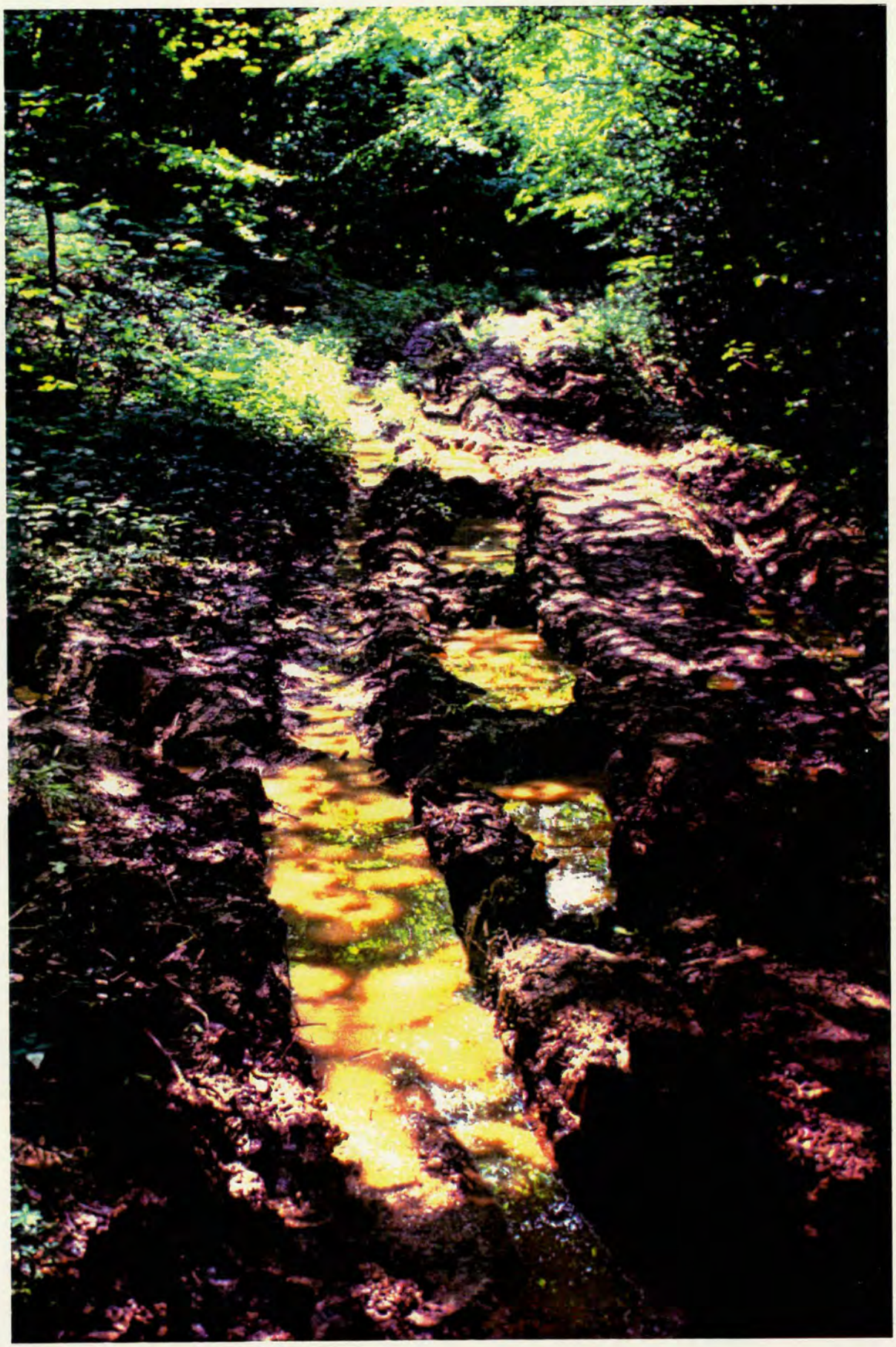


Fig 5.2.3 Puddle enclosures at site 2126. They are the three puddles on the right hand side of the picture.

Retrieving tadpoles

Tadpoles from all the sites were collected once the first individuals were found with extended front and back legs. Two individuals, one from an enclosure in Veleševeć pond and one from site 1039, reached a stage where both front and back legs were extended although the tail had not been resorbed. The contents of all the enclosure bags were collected into separate containers on the 24th June, nineteen days after they were put in. As the bags acted as large sieves this was a relatively simple task. The experimental puddle enclosures were drained the following day (twenty two days from start). Where there was a slope available the puddles were siphoned and tadpoles caught in a small sieve at the end of the hose. Where there was no slope the puddles were emptied bucket by bucket through the sieve. In both cases the last dregs of the puddle were removed and carefully searched. The silt at the bottom of the puddle was also thoroughly checked for tadpoles. Any small movement was easily observed as the tadpoles wriggled vigorously.

The tadpoles were taken to Zagreb University where each individual was anaesthetised, dried on a piece of paper towel and weighed. The snout-vent length was measured with callipers. Length was not recorded for very small individuals as these were kept alive in order to grow them up to provide more tissue for electrophoresis. The presence of limb buds or extended limbs were also noted on a scale of 0-8 (Table 5.2.3). This is a very crude method of measuring development. Normally, more than 40 stages can be distinguished in tadpole development (Gosner, 1960) but time and a powerful microscope are required. The skin colour covering the gut was recorded as it differed between tadpoles being either a shiny coppery orange or a dull mottled grey. Each individual was then frozen separately in an eppendorf tube and stored in liquid nitrogen to transfer to a -70°C freezer in Edinburgh.

Stage	Description
0	No limb buds present
1	One rear limb bud visible
2	Two rear limb buds visible
3	One rear limb bud and one extended rear limb.
4	Two extended rear limbs
5	Two extended rear limbs and one front limb bud
6	Two extended rear limbs and two front limb buds
7	Two extended rear limbs, one extended front limb and one limb bud
8	Both front and hind limbs extended

Table 5.2.3 Description of stages visibly observed with the naked eye in tadpoles collected from the enclosure experiment.

Distinguishing the genotype of different tadpoles

It was thought the gut colour of the tadpoles reflected differences in their genotype. In order to test this, a blind test of tadpoles of known genotype was carried out. Ten tadpoles reared from either *bombina*-like or *variegata*-like parents in the laboratory (by B. Nürnberger) were presented at random to myself and a naive observer. In all cases the tadpoles scored as 'orange' or 'grey' were from *bombina* and *variegata* parents respectively. The colour difference was more rigorously confirmed by scoring the genotypes of a subset of the tadpoles collected from the enclosures (Table 5.2.3 Fig. 5.2.4). Electrophoresis was carried out as described in Chapter 2. The eleven individuals scored as grey were all *variegata*-like with gene frequencies ranging from 0.75 to 1.00. Of the 13 tadpoles that scored orange 11 had very low gene frequencies i.e. they are extremely *bombina*-like. The remaining two individuals scored $\bar{p} = 0.5$ and 1.0. The former mean frequency is averaged across two diagnostic loci only so it may be more *bombina*-like than revealed here. The individual with $\bar{p} = 1$ was scored as orange but it was noted in the original data book that it appeared a mixture of both grey and orange. While the colour coding is clearly not perfect, it nevertheless provides a robust, fast and cheap way of classifying the genotypes of the animals. In any case errors in scoring are likely to decrease any differences observed between genotypes. Unfortunately the colour difference was not observed at the outset of the experiments as the difference became apparent while the measurements in the laboratory were being made. This means that not all tadpoles have a colour score associated with them; tadpole colour was scored at all puddle enclosures and at pond enclosures V2, V3, V6, V8, V9 and V10. It was scored for some of the individuals at V4 but for none at V1, V5 and V7. The way this is dealt with is described below.

Statistical methods

Where appropriate the results were analysed using sign tests. ANOVA was not considered suitable as each taxon within an experimental unit was not independent and the variances between the two groups were not homogeneous. Statistical comparisons between taxa will only include those experimental units where all individuals were colour coded. The effect of habitat on the response of either taxon was assessed by using Wilcoxon two sample tests.

Enclosure (Individual)	No of *variegata* alleles at each locus				$\bar{p}$
	Ak	Mdh	Ldh	Idh	
Grey tadpoles					
v1 (34)	1	1	2	2	0.75
v1 (39)	2	2	2	2	1.00
v1 (41)	1	2	2	1	0.75
v1 (53)	2	1	2	2	0.88
v2 (18)	2	2	2	2	1.00
v6 (16)	2	2	2	-	1.00
v6 (35)	1	2	2	2	0.88
v8 (32)	2	2	2	2	1.00
v8 (35)	2	2	2	2	1.00
v8 (37)	2	2	2	-	1.00
v8 (43)	1	2	2	-	0.83
Orange tadpoles					
v1 (110	0	0	0	0	0.00
v1 (30)	0	0	0	0	0.00
v2 (10)	0	0	-	0	0.00
v2 (11)	0	0	-	-	0.00
v2 (13)	0	0	-	0	0.00
v2 (14)	0	0	-	0	0.00
v2 (27)	0	0		0	0.00
v6 (1)	0	2	-	-	0.50
v6 (17)	0	0	-	-	0.00
v8 (1)	0	0	-	-	0.00
v8 (12)	0	0	-	0	0.00
v8 (3)	0	0	-	-	0.00
v8 (7)	2	2	2	-	1.00

Table 5.2.3 The number of *variegata* alleles scored at each diagnostic locus and the average frequency, $\bar{p}$ across all diagnostic loci, for orange and grey coloured tadpoles retrieved from pond enclosures (see text and Appendix 5.1).

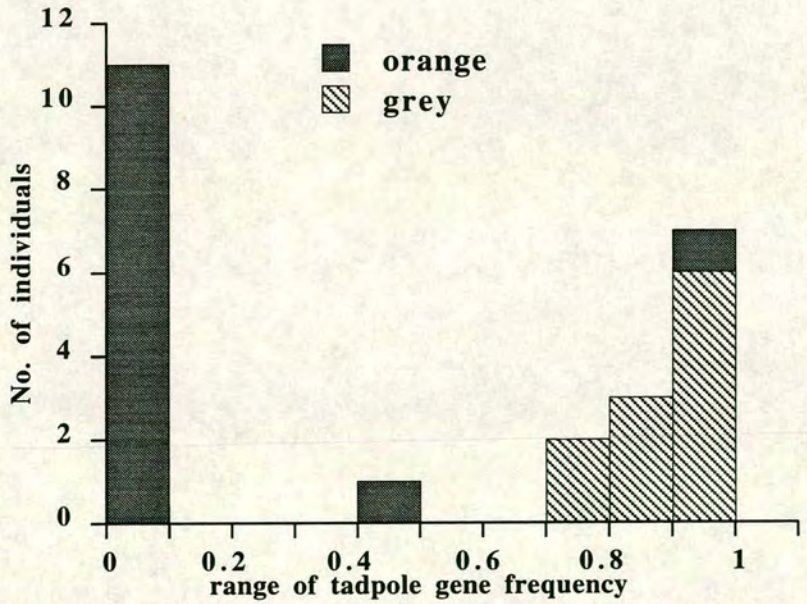


Fig. 5.2.4 Stack histogram showing the range of gene frequencies (measured as $\bar{p}$) of tadpoles scored as either grey or orange. Note that one orange individual has $\bar{p}=1$.

5.3 Results

Individual results of weight, length, stage and colour for each tadpole retrieved in the experiment are given in Appendix 5.1.

Survival

All the pond enclosure bags and five of the puddles contained tadpoles when emptied. There were no tadpoles in the remaining seven puddles. This was most likely due to predation by newts (*Trituris vulgaris*) and salamander (*Salamandra salamandra*) larvae. These were not present in the five puddles which yielded tadpoles but were found in varying numbers at all the other experimental puddles. It was confirmed in the laboratory that newts and salamander larvae ate tadpoles. In puddles survival ranged from 0.00 (where there was predation) to 0.49 (Table 5.3.1). In the pond enclosures survival ranged from 0.10 to 0.54. One pond enclosure (V10, where survival was low; 12%) was found with the neck of the bag twisted and the surface blocked off. As this may have increased mortality and altered development it was excluded from statistical tests. Reasons for variation in survival between experimental units within the same habitat will be discussed below. Overall (excluding the puddles where no tadpoles were retrieved), mean survival between the two habitat types is not significantly different (Wilcoxon two sample test $U_{5,9} = 23$) Table 5.3.1); there is 34% survival in puddles and 31% survival in ponds.

Puddles	No. retrieved	Survival	Ponds	No. retrieved	Survival
P1	41	0.28	V1	78	0.54
P2	61	0.42	V2	27	0.19
P3	27	0.19	V3	51	0.35
P4	43	0.30	V4	59	0.41
P6	70	0.49	V5	15	0.10
P5 and P7-12 had no surviving tadpoles due to predation. These are not included in the means.			V6	49	0.34
			V7	71	0.49
			V8	44	0.31
			V9	36	0.25
			V10	17	0.12
Mean	48.4	0.34		47.7	0.33

Table 5.3.1 Survival rates and number of tadpoles in each of the enclosures. Pond enclosure V10 (shaded) is excluded from the mean and other statistics as the surface was blocked off; this may have altered mortality (see text).

The survival of *variegata* tadpoles was significantly higher than *bombina* (sign test n = 10; p = 0.01; Table 5.3.2, Fig. 5.3.1). This is reflected in the *variegata:bombina* survival ratio which was greater than one in all replicates apart from one pond enclosure (V2). On average the ratio of survival decreased in ponds i.e. *bombina* had higher survival in ponds than in puddles relative to *variegata* . Survival of *variegata*-like tadpoles is on average more than 6 times that of *bombina* in puddles but only 3 times that of *bombina* in ponds (approximately). However the difference is not significant due to the high variance between replicates (Wilcoxon two sample test: $U_{5,5} = 17$; $p > 0.2$). Neither is there a signicant difference in survival between habitats for each taxa separately (Wilcoxon two sample test: $U_{5,5} = 17$ for *bombina* and $U_{5,5} = 18$ for *variegata*; $p > 0.2$).

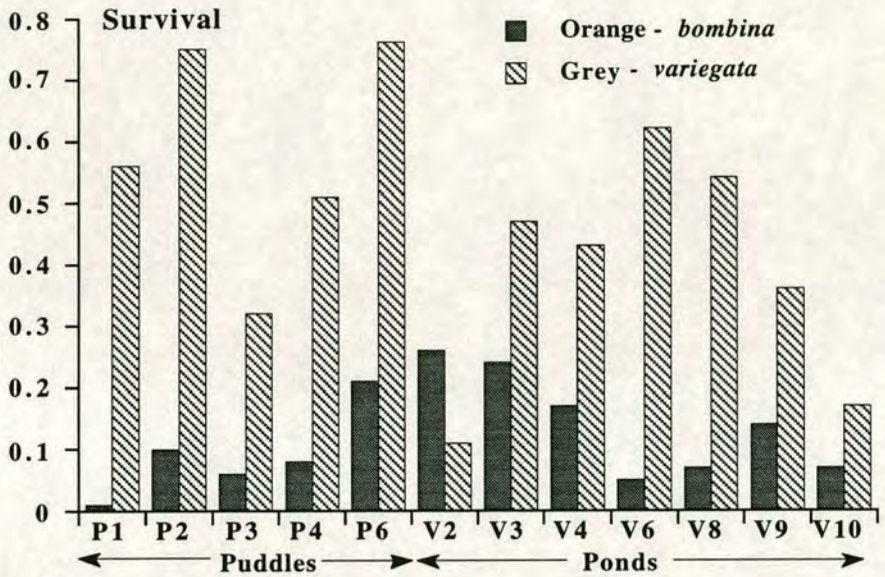


Fig. 5.3.1 Survival rates of each taxon in the experimental enclosures. In all but one pond enclosure (V2) grey tadpoles, i.e. *variegata*-like ones had higher survival rates than the orange coloured *bombina*-like tadpoles. (Survival at V10 may have been altered as the enclosure bag was found twisted with the surface blocked off. As not all individuals were colour coded in V4 these data may also be inaccurate.)

Table 5.3.2 (overleaf) Summary table of survival, weight and stage for *bombina* and *variegata* tadpoles retrieved from puddles or pond enclosure bags. (Only those enclosures where all individuals were colour coded are included; V4 andV10 (shaded) are excluded from means and statistical comparisons; only some individuals in V4 were colour coded while at V10 the neck of the bag was twisted - see text. The difference between the taxa in survival and weight is expressed as the *variegata:bombina* ratio given in the final two columns. Means and standard deviations are given where appropriate. Differences in survival, weight, length and stage are assessed using sign tests and Wilcoxon two sample tests (see text). Overall means are estimated as the average across all individuals rather than the mean of the means.)

habitat	replicate	*variegata*				*bombina*				*v/b ratio*	
		survival (out of 72)	weight mg (s.d)	length mm (s.d)	stage 1-8 (s.d)	survival (out of 72)	weight mg (s.d)	length mm (s.d)	stage 1-8 (s.d)	survival	weight
puddles	P1	0.56	98.0 (32.9)	8.2 (1.03)	0 (0)	0.01	8	-	0	56.00	12.2
	P2	0.75	111.1 (48.0)	9.7 (1.15)	0 (0)	0.10	10.1 (5.6)	-	0 (0)	7.50	11.0
	P3	0.32	127.5 (49.4)	9.1 (1.0)	0 (0)	0.06	13.5 (8.6)	-	0 (0)	5.33	9.4
	P4	0.51	143.7 (84.3)	9.1 (1.64)	0.05 (0.33)	0.08	8.5 (3.4)	-	0 (0)	6.38	16.9
	P6	0.76	145.5 (50.5)	9.3 (1.11)	0 (0)	0.21	14.9 (6.4)	-	0 (0)	3.62	9.8
mean		0.58	125.2 (57.5)	8.9 (1.25)	0.01 (0.14)	0.09	12.4 (6.3)	-	0 (0)	6.44	10.1
ponds	V2	0.11	611.3 (73.6)	16.0 (0.93)	2.75 (1.04)	0.26	234.6 (73.7)	10.6 (1.26)	1.26 (1.19)	0.42	2.6
	V3	0.47	460.3 (129.7)	14.2 (1.65)	2.38 (1.88)	0.24	251.5 (75.7)	11.0 (1.27)	1.53 (0.87)	1.96	1.8
	V6	0.62	482.5 (73.4)	14.5 (0.99)	3.00 (1.69)	0.05	148.0 (30.9)	9.5 (1.00)	1.50 (1.00)	12.40	3.3
	V8	0.54	281.2 (71.0)	12.0 (1.29)	1.69 (0.73)	0.07	126.2 (131.1)	9.3 (2.50)	0.40 (0.89)	7.71	2.2
	V9	0.36	244.9 (41.8)	11.5 (1.07)	0.23 (0.65)	0.14	43.5 (29.2)	7.3 (0.57)	0 (0)	2.57	5.6
	V4	0.43	491.9(104.5)	14.6(1.2)	3.10(1.7)	0.17	240.6(55.4)	10.8(1.0)	1.83(0.6)	2.58	2.04
	V10	0.17	379.5(123.6)	13.3(1.9)	2 (0)	0.07	90.4(31.9)	7.5(1.29)	0 (0)	2.40	4.20
mean		0.42	392.0 (141.45)	13.3 (1.88)	1.78 (1.59)	0.15	188.9 (106.52)	10.3 (1.62)	1.05 (1.08)	2.80	2.1

Morphological measurements and development

Distributions of weight, length and stage are bimodal at almost all sites (Fig. 5.3.3-5). When the colour of each tadpole is superimposed on the distributions (in those enclosures where it was scored) then in almost all cases the orange coloured tadpoles i.e. the *bombina*-like tadpoles are the shorter, lighter less developed individuals (Table 5.3.2, Fig. 5.3.6).

Weight

The weight of *variegata* tadpoles was significantly greater than *bombina* across all replicates (sign test; $p < 0.001$; Table 5.3.2). The difference in weight between the taxa (expressed as the *variegata:bombina* weight ratio; Table 5.3.2), varied significantly between habitats despite the large variation between replicates within each habitat (Wilcoxon two sample test; $U_{5,5} = 25$; $p = 0.005$). On average there was a ten fold difference in weight between *variegata* and *bombina* in puddles whereas the difference in ponds was only twofold. Note that *bombina* showed barely any growth in puddles; the average weight per tadpole was only 12.4 mg compared to 125.2 mg for *variegata*. Taken independently both taxa achieve a significantly greater weight in ponds by the end of the experiment; *bombina* are $\approx$ 15 times heavier in ponds than in puddles and *variegata* are $\approx$ 3 times heavier (Wilcoxon $U_{5,5} = 25$; $p = 0.005$ for each taxon).

Length

Within pond enclosures *variegata* were significantly longer than *bombina* (sign test; $p = 0.031$, Table 5.3.2). Orange tadpoles were not measured for length in the puddles. These tadpoles were extremely small and subsequently reared to increase their tissue mass for electrophoresis. This has two implications, first the fact that all orange individuals could not be scored for length in puddles means that *bombina*-like tadpoles must be shorter on average than *variegata* ones in puddles and second not all *bombina*-like individuals were measured in ponds therefore the length data are for the larger individuals only at some sites, (V8, V9 and V10). Therefore length data in these pond enclosures overestimate the mean (Appendix 5.2 gives individual data).

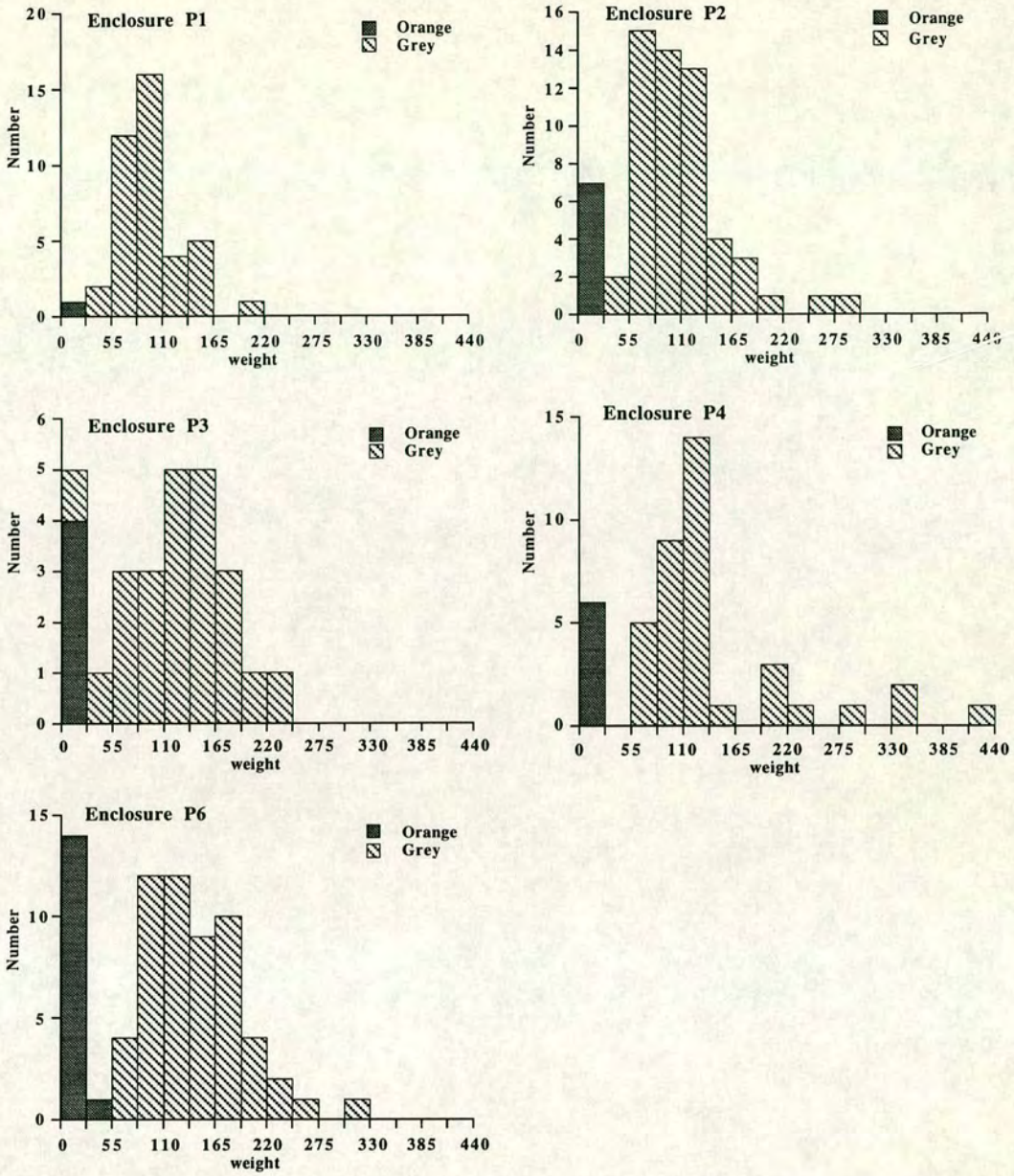


Fig. 5.3.3a Stack histogram distributions of the weight (in mg) of tadpoles from the experimental puddles. The colour scores of the tadpoles are superimposed. Tadpoles scored orange are *bombina*-like and those scored grey are *variegata*-like. The distributions are bimodal. *Bombina*-like tadpoles are lighter than *variegata* (See Table 5.3.2).

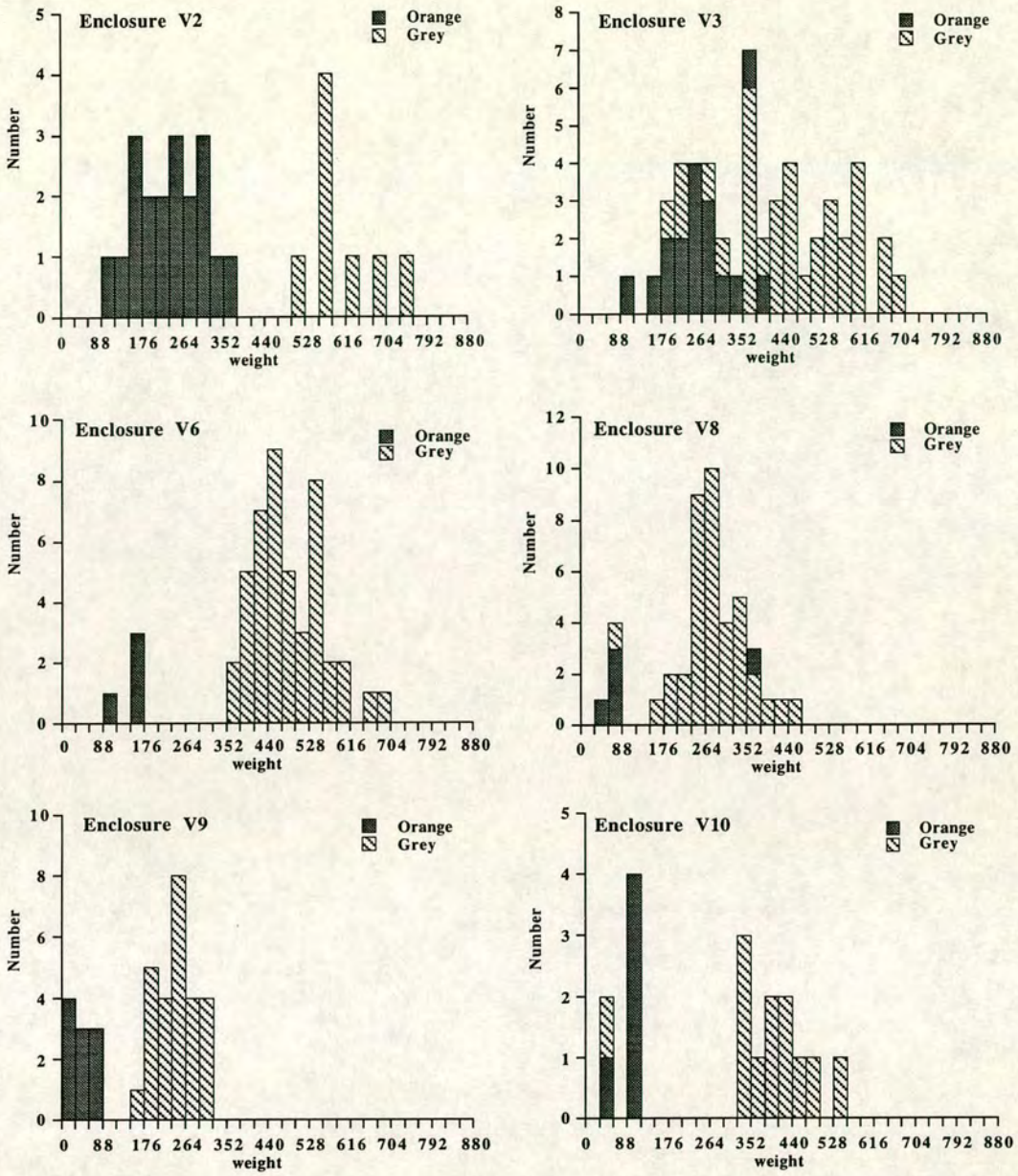


Fig. 5.3.3b Stack histogram distributions of the weight (in mg) of tadpoles from the experimental pond enclosures. Distributions are given for those enclosures where the tadpole colour was scored. The colour scores of the tadpoles are superimposed. Tadpoles scored orange are *bombina*-like and those scored grey are *variegata*-like. The distributions are bimodal. *Bombina*-like tadpoles are lighter than *variegata* (See Table 5.3.2).

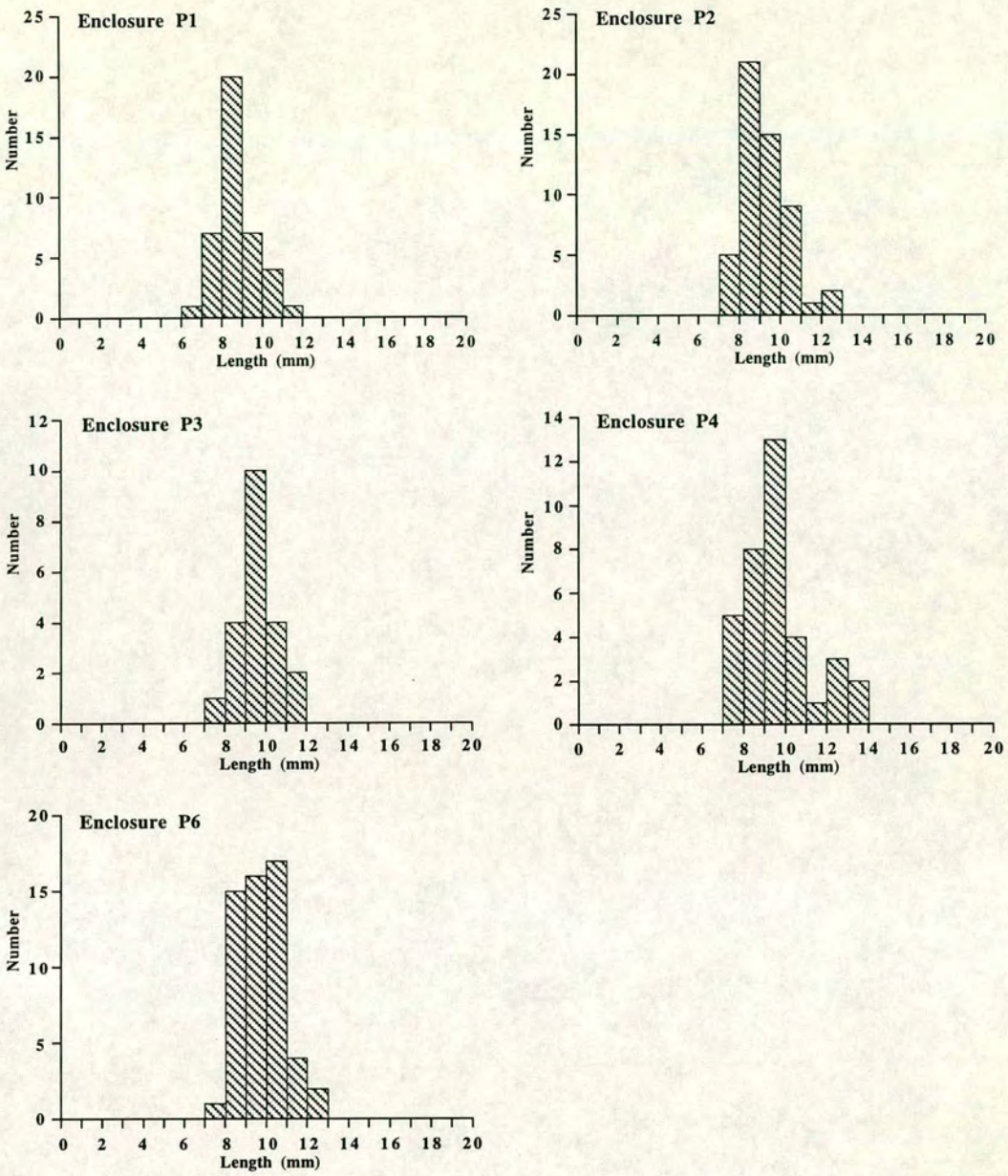


Fig. 5.3.4a Stack histogram distributions of the length of tadpoles from the experimental puddle enclosures. The distributions are for the grey coloured, *variegata*-like tadpoles only (mean values are given in Table 5.3.2). All orange coloured tadpoles were too small to be measured for length (see text)

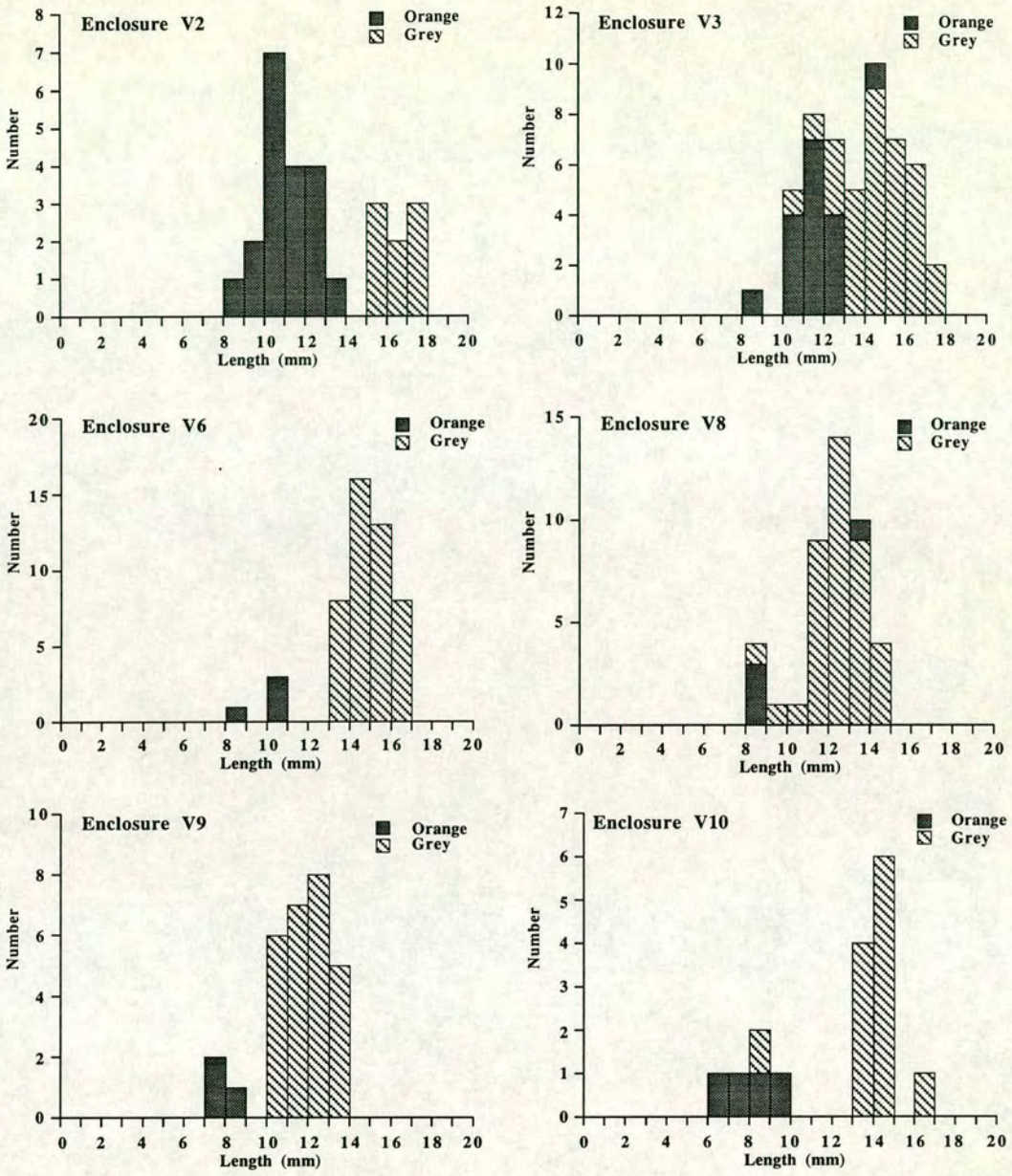


Fig. 5.3.4b Stack histogram distributions of the length of tadpoles from the experimental pond enclosures. The distributions are divided by colour; orange implies that the tadpoles are *bombina*-like while grey are *variegata*-like. *Bombina bombina*-like tadpoles are smaller than *variegata*-like ones (mean values are given in Table 5.3.2).

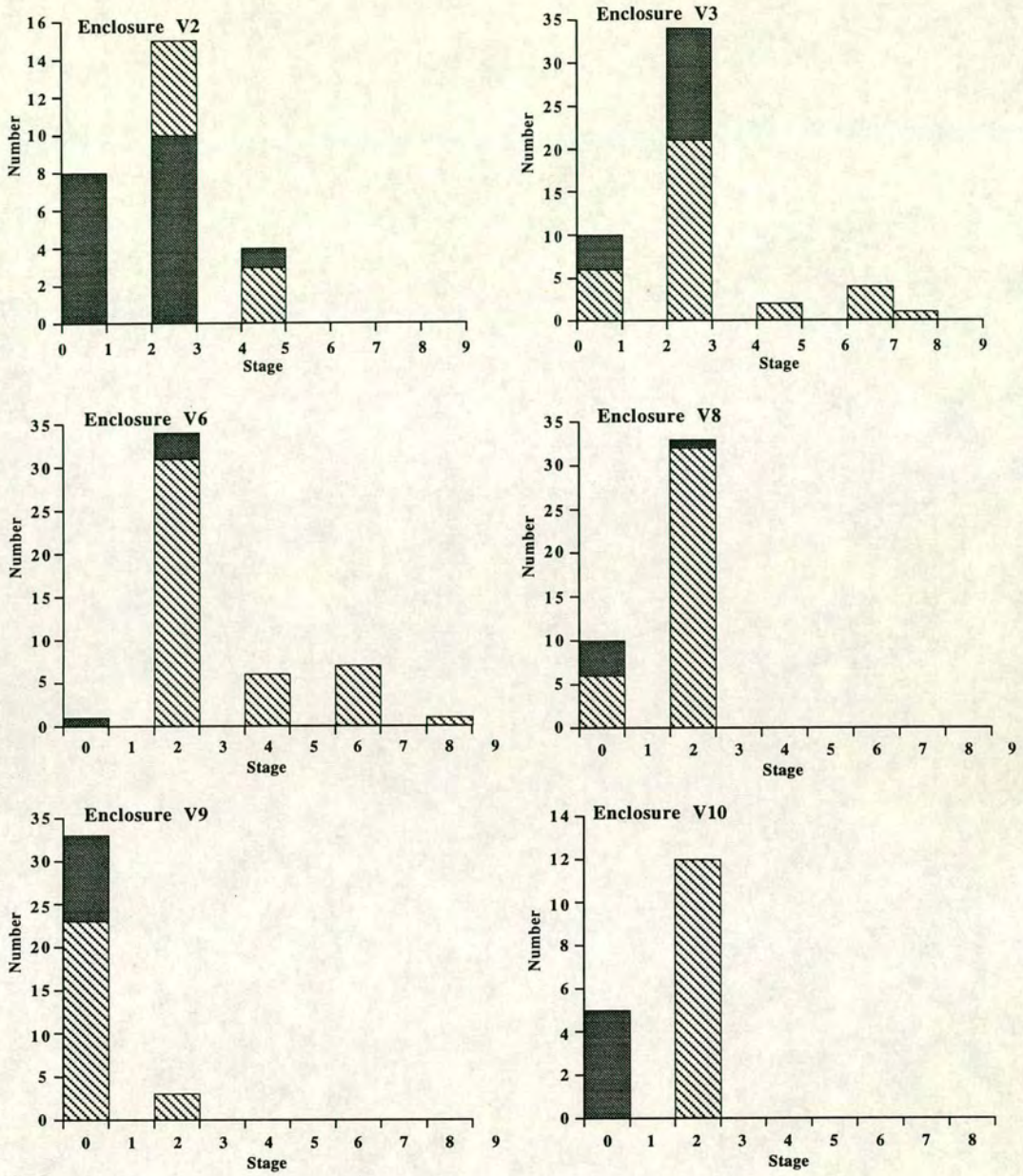


Fig. 5.3.5 Stack histogram distributions of the stage of different tadpoles in each pond enclosure. Stages range from 0-8 (see text for definitions of each stage). The distributions are divided by tadpole colour. The stippled pattern denotes orange while the striped pattern signifies grey. Orange coloured tadpoles, i.e. *bombina*-like ones are at a lower stage than the grey *variegata*-like ones (see Table 5.3.2 for mean values).

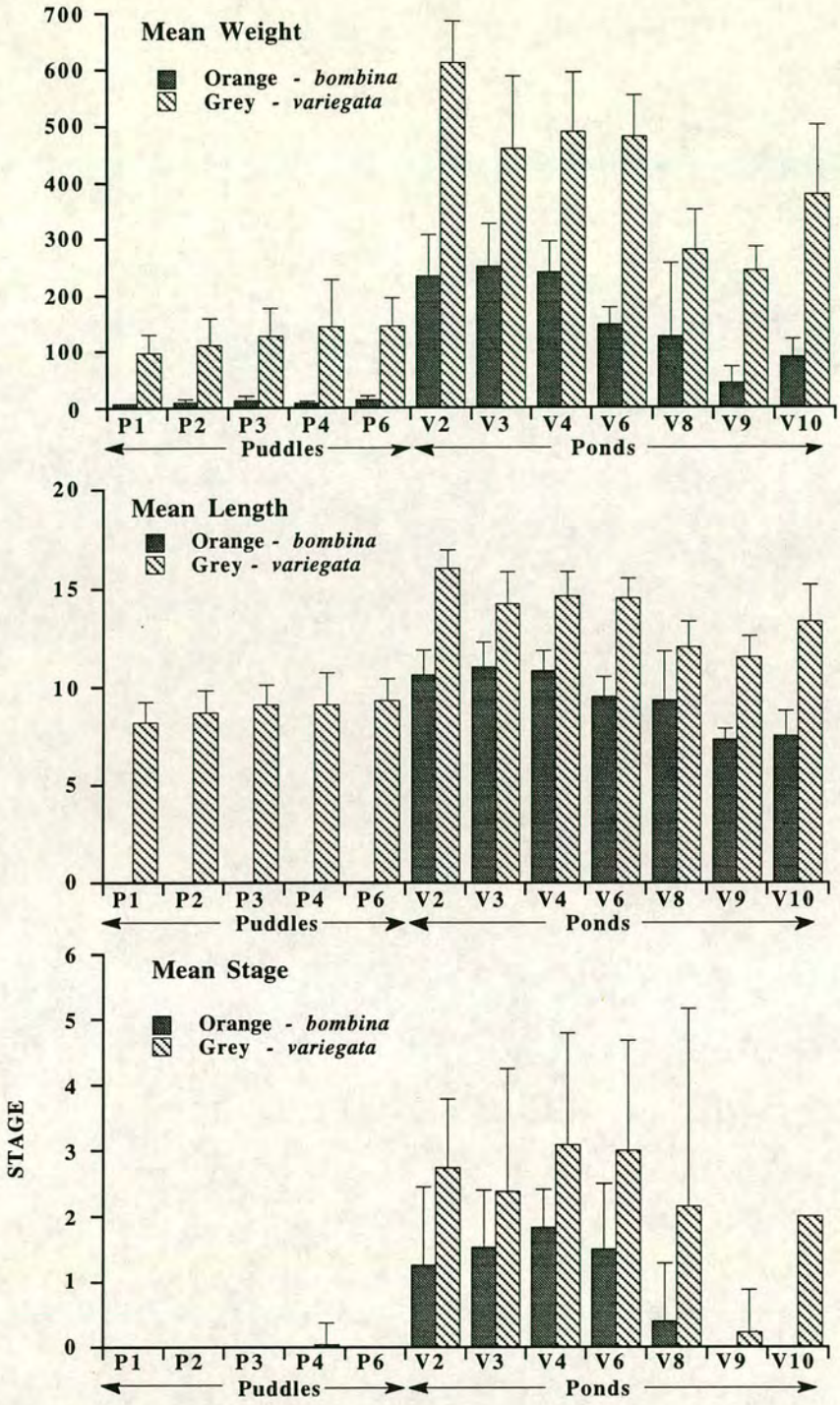


Fig. 5.3.6 Histograms of the means and standard deviations for the weight (mg), length (mm) and stage (1-8) of each colour morph in each enclosure. Orange individuals are *bombina*-like while grey individuals are *variegata*-like. All individuals were weighed and staged. Small individuals were not measured for length. As orange individuals in puddles were extremely small there are no length data associated with them (see text). The stage at most puddles for each colour morph is 0. See Table 5.3.2 for values.

Stage

Individuals in all ponds were at a more advanced stage of development than those in puddles where only a few individuals (in P4) grew beyond stage 0 (Table 5.3.2). In ponds *variegata* were at a significantly later stage than *bombina*. (sign test: p = 0.031). The lack of differences in stage between taxa in puddles is most likely due to the low resolution of the staging. They describe only gross visible differences defined by the presence of limbs. The fact that *variegata* individuals are larger and heavier than *bombina* in puddles suggests that they are in a more advanced stage of development (assuming growth is correlated).

The relationship between weight and length

There is a close correlation between weight and length (Fig. 5.3.7) at all sites. The mean weight of each genotype has been plotted against the mean length and both axes were log transformed. This yields a straight line with a slope of approximately 3 as expected from scaling laws. This implies that as tadpoles get heavier their length increases by the same proportion. The correlation between weight and length is 0.96. In general where there are data on both colour morphs (in the ponds) *variegata* are longer and heavier while *bombina* are shorter and lighter. In puddles there is no length data on the orange tadpoles. The relationship between mean length and weight of *variegata* in puddles and *bombina* in ponds is similar. As length is closely correlated with weight it is assumed that the relationship with genotype and length will be the same as that for weight.

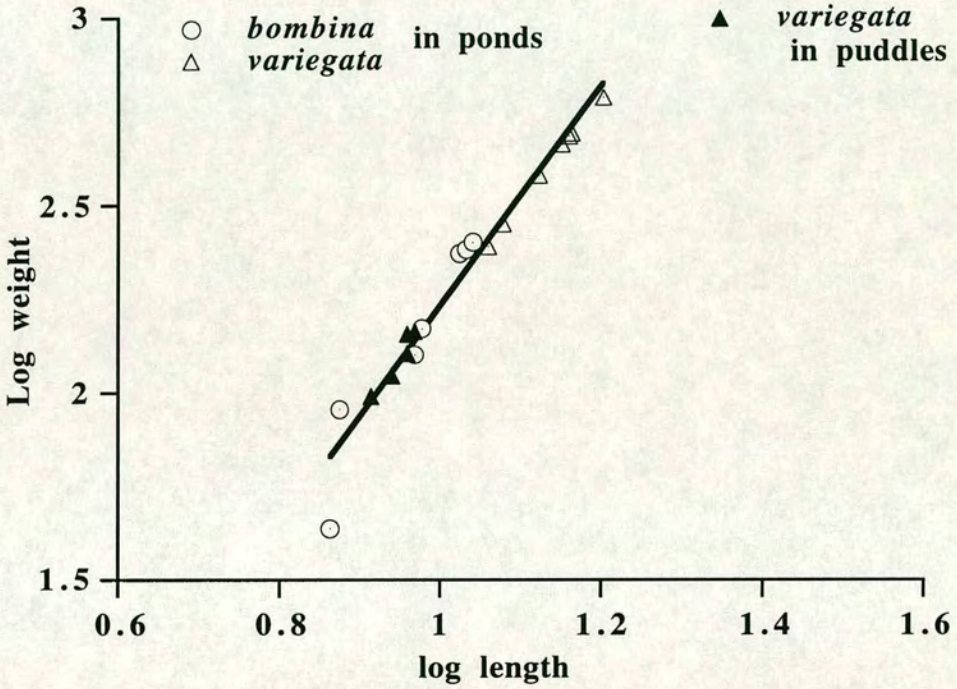


Fig. 5.3.7 The relationship between the mean log length and log weight for orange and grey coloured tadpoles in pond and puddle enclosures. The slope of the line is approximately 3. The correlation of the polynomial function to all the data combined is $r^2 = 0.96$. $(\text{length})^3 \propto$ weight

Sites where colour coding was incomplete

Individuals from the pond enclosures, V1, V3 and V7 have no colour scores associated with them while only some individuals at V4 were colour scored. These enclosures have not been included in the main analysis. However the distributions of weight length and stage can still be examined. At V4 the distributions of weight, length and stage are bimodal (Fig. 5.3.8). It is relatively easy to predict the genotype of those tadpoles of unknown colour by their position in the distribution and strongly suggests that observations around the lower mode represent *bombina* tadpoles.

Individuals at enclosures V1, V3 and V7 were not colour coded at all. The distributions of weight and length at V1 and V7 are also bimodal while those for V3 are unimodal (Fig. 3.5.9). If these enclosures follow the same pattern as the others then the most parsimonious explanation for the bimodality at V1 and V7 is that each mode reflects a different genotype with the mode at the smaller end of the distribution being the *bombina*-like one. V5 does not show a bimodal distribution. The distributions of length and weight appear normal but are at the larger end of the scale for each measurement compared to other distributions where the genotypes are known. This suggests that the *bombina* distributions are absent and only *variegata*-like individuals survived in this enclosure. The distributions of stage at each of these enclosures do not show such an obvious pattern (Fig. 5.3.10). These sites, with no colour coding will not be analysed further.

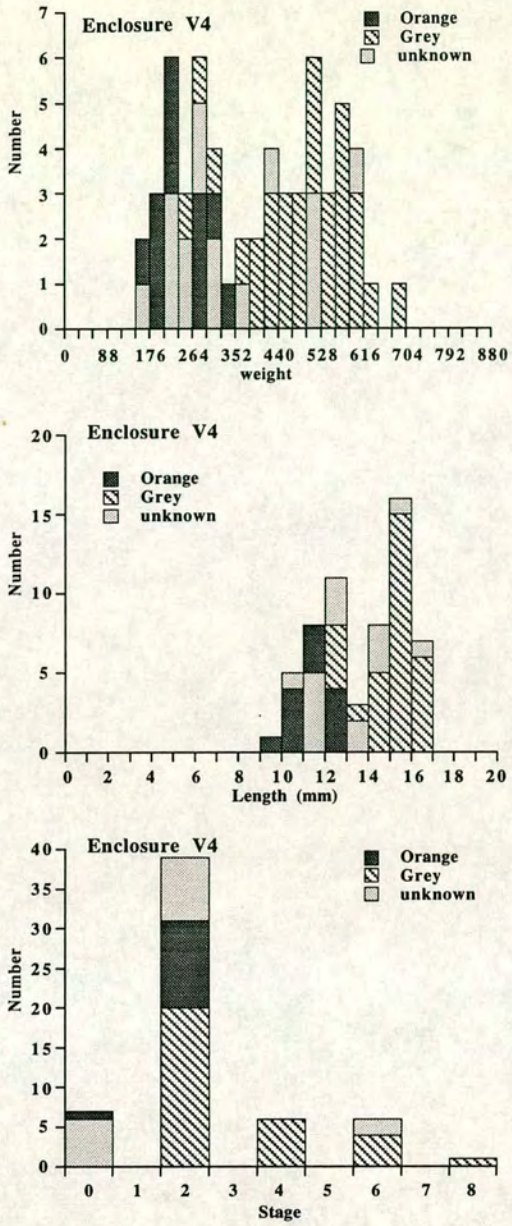


Fig. 5.3.8 Stack histogram distributions of the weight (mg), length (mm) and stage (0-8) of tadpoles in pond enclosure V4. Tadpole colour was scored for most, but not all individuals. The distributions of each measure are similar to those where all tadpoles were scored for colour. The bimodality of the weight and length distributions allows one to predict the colours and hence the genotypes of the individuals whose colour was not recorded. This is more difficult to do with stage (see text).

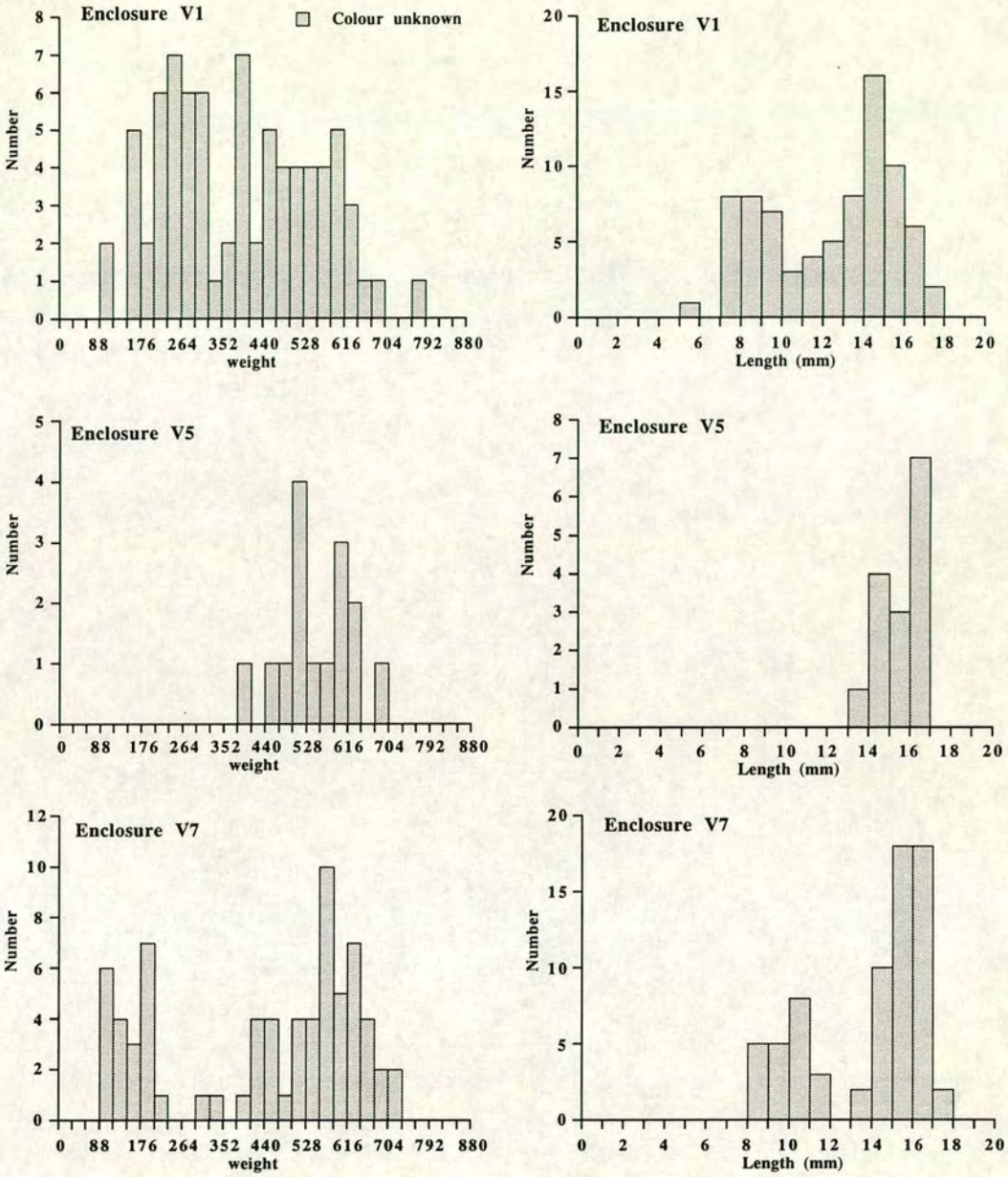


Fig. 5.3.9 Stack histograms of the weight (mg) and length (mm) of tadpoles from those enclosures where the colour of each individual was not scored. The distributions of V1 andV7 are bimodal. If these enclosures follow the same pattern as others then the mode with smaller values are the *bombina*-like tadpoles. The distributions at V5 are not bimodal probably because very few or no *bombina*-like individuals survived at this site (see text).

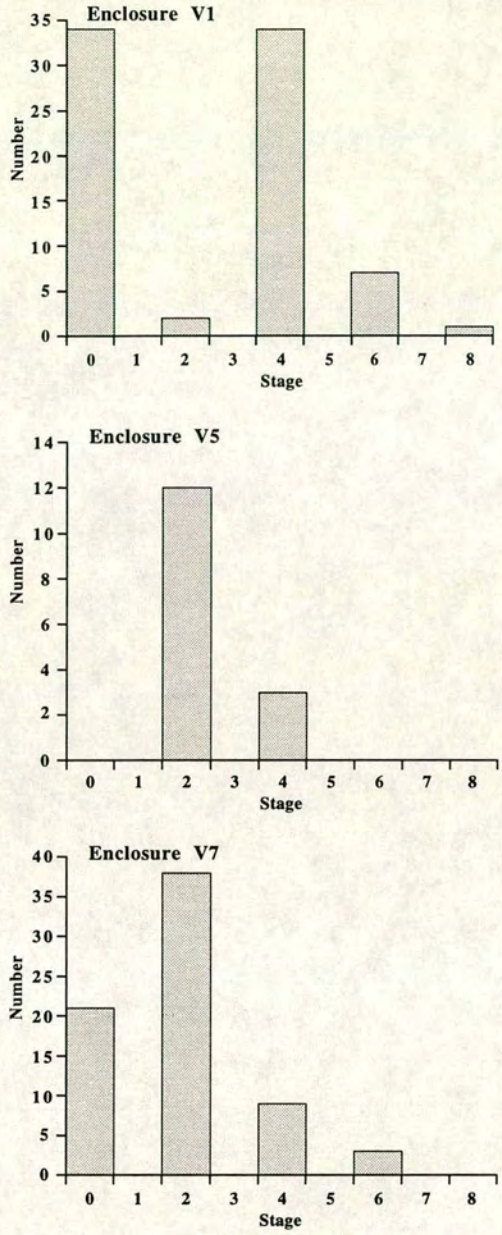


Fig. 5.3.10 Histograms of the distribution of tadpoles at different stages in pond enclosures where individuals were not scored for colour. The stage varies from 0-8 (see text for definition of each stage).

The relationship between egg size and development

The results of differences between genotypes in weight, length and stage in different enclosures demonstrate that overall in both puddles and ponds *variegata* perform better. In both situations *variegata* are longer, heavier and at a more advanced stage. However, the differences between genotypes in puddles is much greater than that in ponds. The ratios of differences observed on average in puddle and pond enclosures are summarised in Table 5.3.3.

Variable	Puddles	Ponds
	variegata:bombina	*variegata:bombina*
survival	6 : 1	3 : 1
weight	10 : 1	2: 1
length	-	1.3 : 1
stage	1 : 1	2 : 1

Table 5.3.3 Ratios of differences between *variegata* and *bombina* tadpoles reared together in puddles and ponds. Ratios are estimated from the overall means for each variable in each enclosure type given in Table 5.3.2.

As data collection for lengths are incomplete, and because staging is crude the ratios for these parameters must be treated with caution. However the relationship between length and weight is highly correlated (Fig. 5.3.7). It is assumed, therefore, that the ratio of $\sqrt[3]{\text{weight}}$ between *variegata* and *bombina* in puddles and ponds will be the same as the ratio of lengths. Despite the problems associated with collecting length measurements, this can be confirmed by comparing the different ratios of weight and length in ponds. The ratio between genotypes in ponds is the same as that for weight (as $\sqrt[3]{2} \approx 1.3$).

At the outset of the experiment the mean egg volume was estimated for eggs collected from puddles and ponds (Table 5.2.2). Egg volume (in conjunction with clutch size) reflects the energy investment of a particular individual (Berven, 1981; Kaplan, 1980a; 1980b). The ratio between egg volume of *variegata:bombina* collected for the enclosure experiment is 4.5. Assuming that the energy content per unit weight is equal in both taxa then *variegata* invest approximately 4.5 times that of *bombina* in an individual egg. Weight ratios of *variegata* in puddles are ten times that of *bombina*: therefore even taking into account the initial difference in investment *variegata* still do better. The difference between genotype weights in ponds however is only two. If

the initial egg investment is taken into account then this means that weight gain of *bombina* is greater than *variegata* in ponds, i.e. they grow by a larger factor in ponds. This is reflected in the fact that *bombina* are 15 times heavier in ponds than they are in puddles by the end of the experiment whereas *variegata* are only 3 times heavier.

The relationship between development and temperature

Differences in weight between *bombina* and *variegata*-like individuals vary between enclosures of the same type as well as between the two types of enclosures. This is especially noticeable in the pond enclosures where for example the mean weight of *variegata*-like individuals ranged from 244.9g (V9) to 611.3g (V2). One reason for this may be that the temperature of the enclosures varied. Maximum and minimum temperatures vary within and between habitats (Table 5.3.4).

Enclosure	Mean Temperature °C			*bombina/variegata* weight ratio
	Maximum	Minimum	Mid	
Puddles				
P1	18.5	13.1	15.8	12.2
P2	18.5	13.1	15.8	11.0
P3	18.5	13.1	15.8	9.4
P4	18.5	13.1	15.8	16.9
P6	16.6	13.5	15.1	9.8
Ponds				
v2	29.0	23.5	26.2	2.6
v3	28.5	19.0	23.7	1.8
v4	28.5	19.0	23.7	2.0
v6	28.0	16.0	22.0	3.3
v8	22.5	12.5	17.5	2.2
v9	22.5	12.5	17.5	5.6
v10	22.5	12.5	17.5	4.2

Table 5.3.4 Range of temperatures observed in the different enclosures. The maximum and minimum temperatures are the mean of a variable number of observations (see Appendix 5.3). The mid temperature is the halfway point between the two. Some enclosures have the same range (e.g.P1-4). This is because all these enclosures are from one site (site 2126) where only one max-min thermometer was situated. The ratio of *variegata* weight to *bombina* weight is greater in the puddle enclosures than in the pond enclosures (see also Figs 5.3.11a and b).

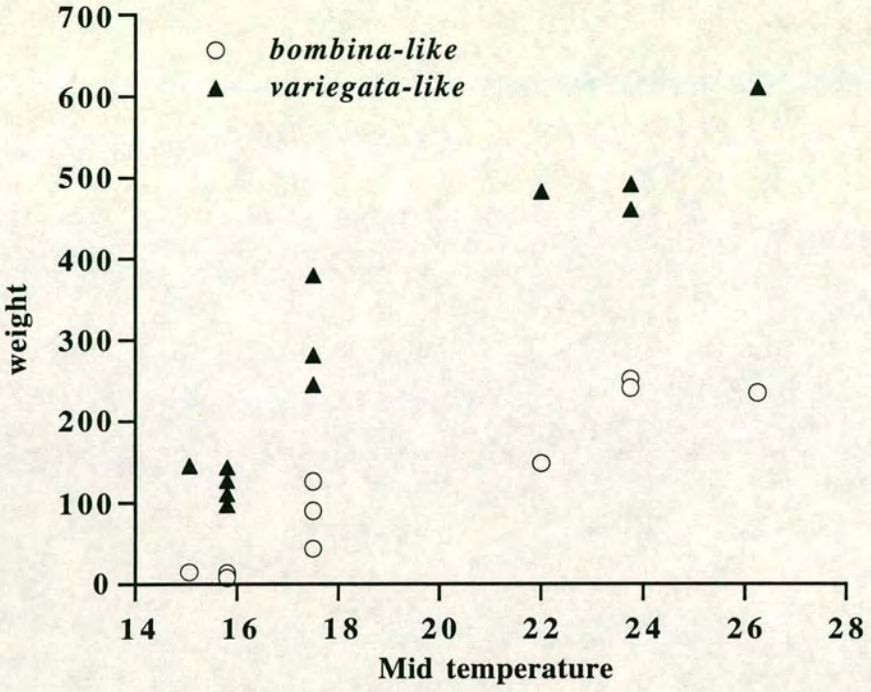
The temperature data are crude. Not all the enclosures contained a maximum-minimum thermometer. For example, at site 2126 where enclosure puddles P1-P4 were situated only enclosure P2 had the thermometer. In this situation the remaining

three puddles at that site were given the same temperature range. Also, as the puddle enclosures were checked more often than the pond enclosures the range of temperatures given are averaged over five occasions at the puddles but only twice at the ponds (see Appendix 5.3 for details of how temperature was estimated for each enclosure).

Despite the drawbacks of the data collection there is a distinct pattern. The observed mid temperatures at the pond enclosures were higher than the puddle enclosures. In all enclosures the weight of *bombina*-like individuals was less than *variegata*-like ones at the same temperature (Fig. 5.3.11a). The weight gain of *variegata*-like individuals is greater over the range of temperatures observed than *bombina*-like individuals. The enclosures varied in temperature from 15-26°C. Over this range *variegata*-like individuals showed a weight difference of on average 513g (98 to 611g) where as over the same temperature range *bombina*-like individuals gained approximately 244g (8 to 252g). However the relative increase in weight is much greater for *bombina*-like individuals where the weight gain is equivalent to a 32 fold increase over this temperature range as opposed to a six fold increase for *variegata*. The ratio of weight differences between the genotypes (within each enclosure) reflect this (Fig. 5.3.11b). As the mid temperature observed in the enclosures increases, then the weight ratio between *variegata* and *bombina* decreases. The differences are dramatic. At around fifteen degrees in the puddle enclosures the weights of *variegata* averaged 9.4 to 16.5 times that of *bombina*. In the pond enclosures where temperature ranged from 17.5 to 26.2°C, *variegata*-like individuals were 1.8 to 5.6 times the weight of *bombina* individuals. The relationship between the weight ratio and enclosure temperature is not linear. There is a sharp drop in the ratio between 15 and 17.5°C. This drop is at the transition between the puddle and pond enclosures.

Temperature is an important parameter controlling amphibian development and activity (reviewed by Berven 1981). At colder temperatures stage specific growth in larvae of ranids is greater resulting in consistently larger sizes during metamorphosis in colder environments. In the wood frog *Rana sylvatica* larger body size of females at reproduction in a colder upland environment results in the production of larger eggs and a greater clutch size than smaller females, (Berven 1981). This is a common phenomenon in amphibians (Duellman and Trueb, 1986).

a.



b.

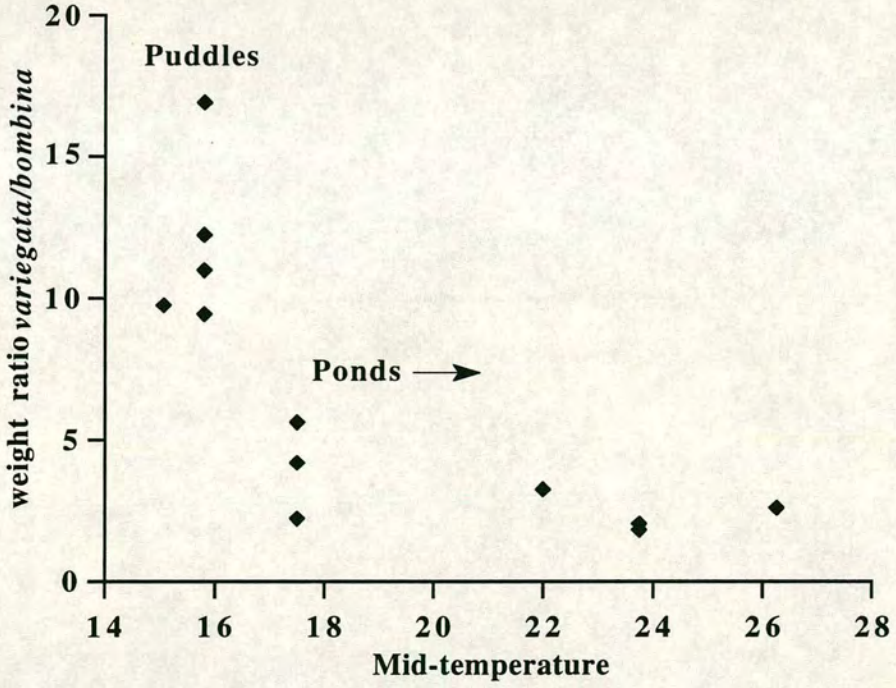


Fig. 5.3.11a,b a)The mean weight (mg) of *bombina* and *variegata*-like individuals in each enclosure given the mid temperature in that enclosure. The mid temperature is the halfway point between the mean minimum and maximum temperatures observed over a variable number of occasions (see Appendix 5.2). The genotypes of individuals in each enclosure were identified by their colour (see text for details). b) The ratio of the mean *variegata* to *bombina* weight in each enclosure as a function of its temperature.

Egg size affects the survival of embryos at colder temperatures. Larger eggs use a smaller proportion of the total energy content to hatch resulting in hatchlings with a higher energy content (Kaplan 1980a). In fish it has been shown that when food is a limiting factor then larvae from larger eggs had more energy reserves in their yolk sac than those from smaller eggs and were able to survive longer at colder temperatures (Blaxter and Hempel, 1963). Berven (1986) has demonstrated the low survival of small lowland eggs compared to large ones at cold temperatures in *Rana*. This may well account for the low survival of *bombina* in the puddle enclosures. Both *variegata* and *bombina* were released into puddles as hatchlings still containing their yolk sacs. In puddles only 9% of *bombina* survived compared to 58% of *variegata*. However in one puddle, P6, *B.bombina* had much higher survival than other puddles despite it being the coldest (Table 5.3.2 and Table 5.3.4). The reason for this may be that food availability at this puddle was greater than at the other puddle enclosures. P1-4 were fresh wheel ruts with a substrate of mud. P6 however had a great deal of leaf litter on the bottom. *Bombina* larvae graze on algae which may have been in greater abundance at this site. This would not only account for the higher survival of *B.bombina* at this enclosure but also their greater weight.

One of the arguments against the translocation experiment done here is that eggs laid by females reared in a hotter environment i.e. *bombina* would lay smaller eggs than the same individuals reared in a colder environment. In other words the fitness difference observed here might be eliminated were pure *bombina* raised in an upland environment where they would attain a larger body size and hence lay larger eggs. However Berven (1981) has shown that in the wood frog egg size and number are not as plastic to environmental conditions as body size. He translocated juvenile frogs between an upland and lowland environment and compared body size, age of first reproduction, egg size and clutch size with the resident population in each place. The transplanted individuals matured at an age and size intermediate to that of the resident population and the population of origin. This demonstrated that both these traits had a genetic and an environmental component. However by allowing for differences in body size, Berven showed that egg size and egg number were solely determined by the population of origin. There is an environmental effect on egg size due to those effects on body size rather than directly affecting eggs. This has important implications for this study as *B. bombina-like* individuals dispersing from a lowland to an upland environment will be at a disadvantage in two respects Their egg size will

be smaller due to genetic constraints and due to a body size determined in the lowlands.

If egg size responded to differences in temperature directly then the fitness difference in opposing habitats may not be as great as the results of this experiment suggest. Kaplan (1987) has shown that female *Bombina orientalis* respond to an increase in nutrition or a decrease in temperature by laying larger eggs and that this was independent of body size. This result directly contradicts Berven's above. It may be that environmental response varies between species. There is some evidence that egg size may vary with temperature at least for *B. variegata*. Eggs laid by *variegata* under laboratory conditions are significantly smaller than those collected from the field (Nürnberger *et al.*, 1994). At a constant temperature of 22.5°C the mean egg volume of *variegata* is 4.54mm^3 (s.d = 0.77) compared to 7.79 (s.d = 1.55) in the field. There is little difference in the egg volume of *bombina*, 1.96mm^3 compared to 1.71mm^3 in the field. This may be because laboratory temperatures are closer to the temperatures recorded in the ponds (Table 5.3.4). However it is difficult to separate temperature from nutritional effects. Poor nutrition also results in a smaller egg size for many amphibians (Berven 1981). Laboratory food may not meet the nutritional requirements of the animals. But this does not explain the lack of difference between the egg volume of *B.bombina* between field and laboratory. However it has been shown that the nutritional status of *bombina* females has less effect on egg size than it does for *variegata* (Nürnberger, unpubl.).

Limitations and conclusions of the experiment

There is no doubt that under the conditions of this experiment *variegata* survive better and grow more quickly than *bombina* in puddles but given the difference in egg size *bombina* gain weight at a greater rate than *variegata* in pond enclosures. Taken independently both taxa achieve a greater weight in ponds. Overall this suggests differential adaptation to habitat.

There are a number of criticisms that can be directed at this experiment. The most obvious is that while the puddle enclosures may be representative of upland puddle habitat, bags suspended in ponds may not represent pond habitat at all. Within a bag all flora and fauna are largely excluded. This will not only eliminate potential predators but may also limit the foraging opportunities for the tadpoles. This latter

consideration turned out not to be a problem as there was extensive algal growth on the side of the bags on which the tadpoles used to graze. However, this introduced an alternative problem of artificially increasing the food supply or at the very least changing it from the norm. It is likely that food availability differs anyway between puddles and pond. It is hard to identify what tadpoles might feed on in puddles whereas ponds have a rich variety of vegetation.

The exclusion of predation must also have had a significant effect. Predation at the puddle sites eliminated seven enclosures from the experiment. It is likely that predation in a species rich pond is equally important. It is known that *bombina* lay a larger quantity of smaller eggs (Nürnberger *et al.* , 1994; Rafinska, 1991). One reason for this may be to reduce the effects of predation on a single clutch (Berven 1981). As *variegata* lay fewer eggs then predation in a pond may result in relatively greater mortality for a *variegata*-like clutch than a *bombina* one. Also *bombina* tadpoles may have a better behavioural strategy to limit predation.

Another problem is that no consideration of density was made in either enclosure type. The puddles varied in size and while the pond bags themselves did not, the depth of water they were suspended in did. The number of individuals put in each enclosure was determined by the number of eggs collected. As numerous eggs are laid in both puddles and ponds it was assumed that density in the enclosures would be less than that in reality, though this was never tested. It is well known that high densities affect larval development (e.g. Travis 1984). Density may also have different effects on different genotypes.

A decision to put equal quantities of the two genotypes in each enclosure together, was designed to reveal relative differences between them. If different genotypes had been placed in separate enclosures then it would be difficult to disentangle differences between genotypes from differences between enclosures. However this immediately sets up the artificial situation of putting pure types in competition with each other. It is hardly likely that such a combination of genotypes would exist in the centre of the hybrid zone. Also genotypes transported to the alternative side of the zone were not only placed in different aquatic habitat types but at different altitudes and surrounded by a different habitat. It would have been ideal to control for this by putting the same combination of genotypes in a different aquatic habitat but on the same side of the zone.

The final criticism is that this experiment reveals what happens to eggs laid in lowland and upland habitats but reared in the opposing habitats. This eliminates all the environmental contribution to egg size and body size of adults that the resident population have. Ideally one would wish to do a similar experiment to Berven (1986) and rear a range of juveniles genotypes in upland and lowland areas and compare the resulting adult size and size and development of their offspring. This would reveal fitness differences in the pure habitats.

It would be interesting to know what the habitat differences are within the hybrid zone. Here puddles may be more exposed and have higher temperatures; puddle habitat may not therefore be as hostile an environment to *bombina* in lowlands as upland puddles are. Also what are the relative fitnesses of hybrid populations in various habitats across the zone. Mean egg volume for hybrids is intermediate between that for the pure populations. It may be that they are relatively fitter than *variegata* in ponds or *bombina* in puddles if selection is due to adaptation rather than any internal genetic incompatibilities within hybrids.

Despite the limitations of this experiment it can be concluded that there is definitely selection in relation to the environment. *B. variegata* are fitter than *B.bombina* in their own habitat type in terms of survival and weight gain. Whether *bombina* are fitter than *variegata* in ponds can not be concluded on the basis of this experiment though taken independently *bombina* are better adapted to a pond environment than a puddle one. Given the difference in egg size (and the assumption of equal investment) then *bombina* grows by a larger factor than *variegata* in pond enclosure bags. This trend however is reversed for survival. As predation is excluded in pond bags genuine survival estimates can not be determined. Despite the caveats these data strongly suggest that the habitat preference expressed by the adults is adaptive.

Synopsis of results

1. The total number of individuals surviving from puddle or pond enclosures was similar. Overall survival in puddle enclosures was 31% and in ponds it was 34%.

2. Survival differed significantly between genotypes within all enclosures. In all experimental units except one *variegata* had a higher survival than *bombina*.. Although the total difference in survival between genotypes from puddles was greater than from ponds the variation within puddles and ponds was large and the difference was not significant.

4. *B.variegata* were significantly heavier than *bombina* in both puddles and ponds after three weeks. Taken indendently both taxa were significantly heavier in ponds than they were in puddles. Where data are available *variegata* were significantly longer and generally at a more advanced stage of development than *bombina*.

5. The differences in weight between *bombina* and *variegata* were significantly less in ponds than in puddles. Overall *variegata* were approximately ten times the weight of *bombina* in puddles but only twice their weight in ponds.

6. The average volume of *B. variegata* eggs collected for this experiment was 4.5 times that of *B. bombina*. Taking this initial investment into account means that *variegata* did better than *bombina* in puddles but *bombina* performed better than *variegata* in ponds (in terms of relative weight gain).

7. There was a relationship between the difference in weight of the two genotypes and the temperature of the enclosure in which they were reared. In general the pond enclosures had a higher observed temperature than the puddle enclosures. The absolute gain of weight was greater for *variegata* than *bombina* as temperature increased however the *variegata/bombina* weight ratio decreased from a maximum of 16 to a minimum of 2 over the same range of temperatures.

9. Differences in temperature and food availability between sites may account for the variability in performance seen within puddle and pond enclosures.

Chapter 6

The role of environmental heterogeneity in the *Bombina* hybrid zone.

A summary of the main results from this thesis is given in Table 6.1. *Bombina bombina* and *Bombina variegata* are found in different habitat types in Croatia. In general *bombina* is found in ponds in more arable regions while *variegata* is found in puddles in upland forests. They hybridise where they meet at the transition between lowland and upland. Populations with a wide variety of genotypes have been described in the centre of the zone. Their distribution can be described by a stepped cline where there is a difference in gene frequency between different habitats. Evidence has been provided for a habitat preference; despite movement between sites the relative difference in gene frequency between different habitat types remained consistent though both were available to a range of genotypes. It has also been demonstrated that the different taxa are adapted to their own habitat. These results differ from analyses of the *Bombina* hybrid zone in Poland. The Polish cline shows a sharper step in the centre with a smoother transition of genotypes across the cline (Fig. 6.1). The cline at Pešćenica shows much more fluctuation from place to place, in part due to habitat preference.

The following discussion is divided into a number of sections. Section 6.1 will outline inferences that can be made from the strength of disequilibrium and the shape of the cline based on the same assumptions used to make inferences from the transects between *bombina* and *variegata* in Poland (Szymura and Barton, 1986; 1991; Chapter 1). These inferences do not take into account the adaptation to habitat and habitat preference that are observed here. Sections 6.2 and 6.3 will explain the observed differences in the distribution of genotypes and the shape of the cline between Poland and Pešćenica in the light of the preference and adaptation to habitat.

Table 6.1 Summary of main results in present study

Chapter 2 - Pattern and distribution of genotypes

Distribution of gene frequencies	There are few sites whose populations have an intermediate gene frequency.
$F_{st'} = 0.00681$	Variance in gene frequency estimated from discordance between loci
$F_{is} = 0.26$ in centre $R = 0.388$ in centre	Deviations from Hardy-Weinberg Linkage disequilibrium (significant at edges) Fis and R stronger on *bombina* side of zone
Cline Length = 36km $L_{123} = -251.96$	147 sites Described by 9 segments each 4km long Significant and large residual variation

Chapter 3 - Quantifying a difference in gene frequency between habitats

Habitat type

	Aquatic	Terrestrial
variegata-like	puddles (little vegetation)	forest (upland)
bombina-like	ponds (vegetation present)	arable (lowland)

Δp (H) = 0.5	Observed difference in gene frequency between aquatic habitat types in the centre of the zone (centre is where the mean gene frequency across both habitat types is 0.5)

Determining the position and shape of the cline incorporating a habitat difference.

$p(\text{exp}) = p + \alpha Hpq$	The expected gene frequency of a population in one of two habitats (H = -1 or 1) given that the difference in gene frequency will vary as α.
N = 117	Number of sites in analysis (only those with a defined habitat type were used)

a) In two dimensions

Constant width and α width = 4.67km (2.32, 5.13) $\alpha = 0.37$ (0.28, 0.48)	$L_{98} = -176.136$ (likelihood of cline)
Variable width and α width = 0.06km - 6.73km $\alpha = 0.23$ (0.17, 0.37)	$L_{98} = -145.59$ width varies significantly

Table 6.1 continued

b). Determining the position and shape of the cline incorporating a habitat difference in one dimension.

N = 134	Number of sites in analysis
L_{126} = -118.49	Likelihood of cline
y = 0.01 (0.05, 0.09)	Centre of the cline (distance standardised by the width)
w' = 1.08 (0.88, 1.39)	The standardised width. The position of each site from the centre of the cline is standardised by the width of the cline at that point and then reduced to one dimension. Therefore both the position of the centre of the cline (y) and the width (w') have no units associated with them.
α = 0.31 (0.13, 0.42)	Estimated difference in gene frequency according to habitat. (In the centre of the zone the diagnostic gene frequency differs between habitat types by $\alpha/2$).
Fst = 0.025 (0.02, 0.05)	Variance of concordant fluctuations across loci between sites.
θ_b = 0 (0-0.06) θ_v = 0.07 (0.03, 0.19)	Measures of the rate of decay of the tails of introgression into *bombina* and *variegata* respectively; decay described by $\exp(-x\sqrt{\theta}/w)$
B_b/w = 40.9 (6.8, 666.2) B_v/w = 3.3 (1.5, 8.1)	Measures of the barrier (in widths) to gene flow into *bombina* and *variegata* respectively. $B=\Delta p/(dp/dx)$ where Δp is the change in allele frequency across the step.
r = 0.25	Harmonic mean recombination rate between enzyme marker and selected loci (from Szymura and Barton 1991)

Inferences

σ = 1.03 km $gen^{1/2}$	Dispersal rate (defined as the standard deviation in distance between parent and offspring.
s* =0.387 (0.095, 0.468)	The effective selection pressure on an enzyme marker (defined by $w = \sigma^2/s^*$)
s_e (b) = 0.027 (0.012, 0.074) se (v) = 0 (0, 0.023)	The selection acting directly on a *bombina* or *variegata* allele outside the hybrid zone. (estimated from $\theta = se/s^*$)

Table 6.1 continued

Chapter 4	**Evidence for dispersal between sites and a habitat preference.**
r = 0.61	Overall recapture rate
358m (N = 20)	Average distance observed of individuals that moved during a field season.
929m (N = 7)	Average distance moved within a year (of individuals that moved).

Despite movement between sites and a high variance in gene frequency the mean gene frequency within a site remained constant.

Chapter 5	**Evidence for adaptation in relation to habitat in the enclosure experiment.**
***bombina : variegata* ratios**	
1: 4.5	Egg volume from ponds and puddles; can be interpreted as initial energy investment.
In puddles enclosures	Averaged over 6 experimental puddles
1 : 6	survival
1 : 10	weight
1: 1	stage
In pond enclosures	Averaged over 7 pond enclosure bags
1 : 3	survival
1 : 2	weight
1: 2	stage

Both *variegata* and *bombina* are significantly heavier in ponds than in puddles. On average *bombina* are fifteen times heavier in ponds than they are in puddles whereas *variegata* are only three times heavier.

Differences in weight between ponds and puddles may be related to temperature.

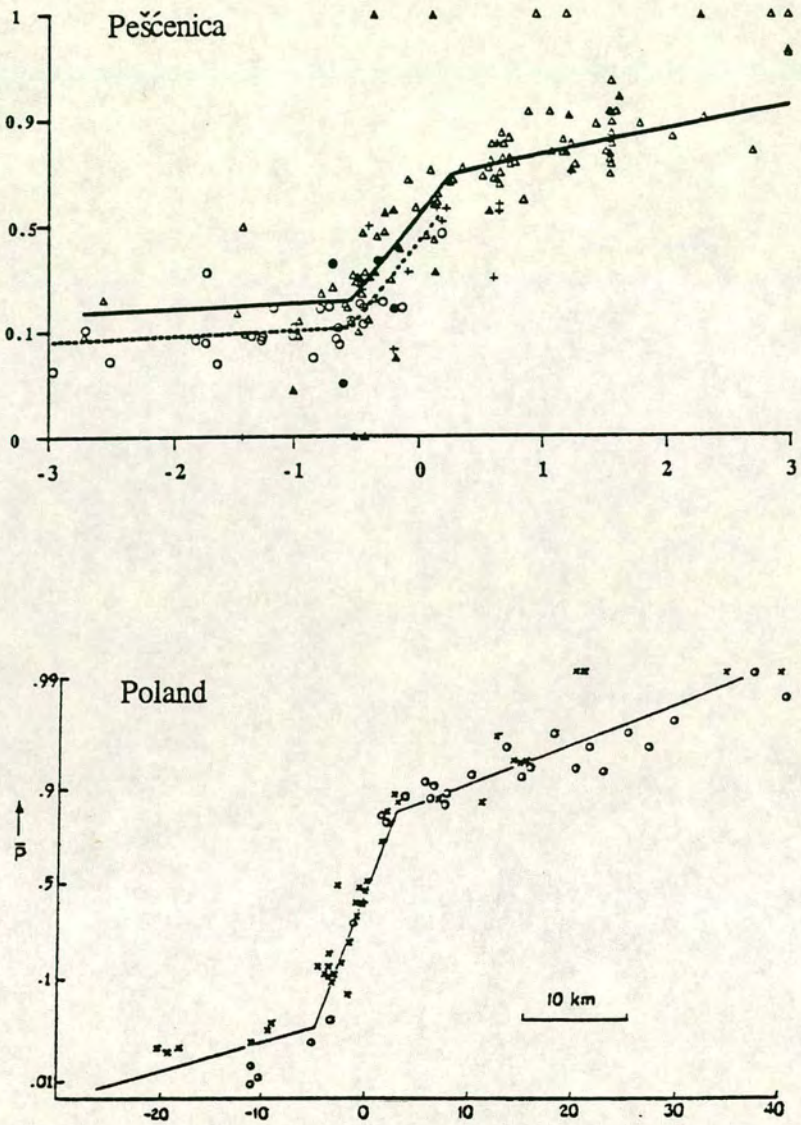


Fig 6.1 The pattern of gene frequencies across the transect in this study (top, from Chapter 3) and across the two Polish transects at Cracow and Przemyśl (bottom, from Szymura and Barton, 1991). The mean *variegata* frequencies of both graphs are plotted on a logit scale ($\log_e[p/q]$). The range of frequencies is from 0-1 at Peščenica as some populations are fixed. Note the difference in gene frequency between different habitats at Peščenica reflected by the double cline; one through ponds (truncated line) and one through puddles. Note also that the step in gene frequency at the centre of the cline is larger in Poland.

The reasons for the difference between the transects will be discussed in section 6.4. Finally the implications of the results of this thesis for the mechanisms of speciation will be assessed in section 5.

6.1 Inferences from disequilibrium and the shape of the cline

The cline at Pešćenica shows significant variation in width along its length (Chapter 3). In order to make inferences from the shape of the cline the results from the one dimensional fit will be used (Chapter 3.4). Selection acting at one locus should result in a smooth sigmoid cline which appears as a straight line on a logit scale (Bazykin 1969, Barton 1979). The cline here shows a step in the centre which reflects a barrier to gene flow across the zone. This is to be expected given the strong disequilibrium between the marker loci demonstrated in the Chapter 2. The step reflects the effect of selection on the whole genome which acts on any locus via linkage disequilibrium (Barton 1983, Barton and Gale 1993).

Estimating the dispersal rate

Assuming that the strong disequilibrium in the centre of the zone is maintained by the dispersal of parental genotypes into the zone, an indirect estimate of the dispersal rate ' σ' (measured as the variance in parent-offspring distance) can be made (Chapter 1). Disequilibrium is estimated across all pairs of loci (where the likelihood is summed across pairs), and standardised by the mean gene frequencies ($R=D/\sqrt{pquv}$). Substituting R for D in equation 1.3 of Chapter 1 gives:-

$$R = \sigma^2 p'^2 / pqr = 4\sigma^2 / w^2 r$$

(Szymura and Barton 1986) where p' is the gradient of the cline, w is its width and r is the recombination rate between unlinked loci in the centre of the hybrid zone (=0.5). Therefore the relationship between dispersal rate and the strength of disequilibrium measured as R is

$$\sigma = w\sqrt{Rr/4} \qquad (6.1)$$

R was estimated as 0.388 at its peak (Chapter 2). Given that the average width of the cline is 4.67km (limits 2.32 and 5.13); when the cline width is constrained to be

constant), then dispersal is estimated as 1.03 km per generation (0.51, 1.13km). This estimate is close to that for the Polish transects where σ is 0.99 and 0.89km.gen$^{1/2}$ for Przemyśl and Cracow respectively.

Estimating the effective selection against a marker locus

Again, assuming that the cline is maintained by the dispersal of parental genotypes into the zone and selection against hybrids within it then the width of the cline is proportional to the ratio between dispersal and the square root of selection. For the specific model of selection against heterozygotes and weak selection (i.e.<0.1) the width is $w = \sqrt{8\sigma^2/s^*}$ (equ 1.4 Chapter 1). The estimates of dispersal and average width mean that the 'effective' selection against recombinants in this hybrid zone is s* =0.388; note that this is equal to the standardised disequilibrium R.
This can be shown as follows. Substituting for width in equation 6.1 gives

$$\sigma = \sqrt{8\sigma^2/s^*}.\sqrt{Rr/4}$$

therefore:-

$$\sigma^2 = \frac{8\sigma^2}{s^*}.\frac{Rr}{4}$$

and as r, the recombination rate is 0.5 then s*=R.

This calculation implies that selection of 39% is required to maintain the strong linkage disequilibrium. This selection is not that on an individual locus but is a reflection of the selection on all the selected loci with which it is in disequilibrium. Associations will break down as a gene crosses the hybrid zone. Therefore at the edges of the zone one would expect both disequilibrium and the effective selection acting on a particular locus to be weaker. Linkage disequilibrium is indeed weaker though still significant at the edges (Chapter 2). This could be due to a few individuals moving long distances (Szymura and Barton 1986); other reasons for an inflated disequilibrium will be discussed below. The effective selection estimated to act on a marker in the Polish transects is 0.22 at Przemyśl and 0.17 at Cracow. The effective selection here is approximately twice the strength.

The shape of the cline

The step in gene frequency at the centre of the cline can be interpreted as a barrier to gene flow. It is primarily a description of the shape of the cline. Inferences can be made about the barrier to gene flow and the rates of introgression from differences between the step in the centre and the slope of the tails on either side of the zone (Chapter 1). It is difficult to make inferences from the barrier strengths and rates of introgression as the relative contributions of each to the shape of the tail are unclear (Chapter 3). The estimates are more certain on the *variegata* side of the zone. Here the rate of decay of introgressing alleles is θ_v = 0.07 (limits 0.03, 0.19) given that F_{st}= 0.025. Since θ can be interpreted as the ratio between the selection on an introgressing enzyme marker in the tail (s_e; selection acting on enzyme) and the effective selection on the marker within the hybrid zone (s_e/s^*), this implies that the selection acting against a *bombina* allele on the *variegata* side of the zone = 0.027 (0.0010-0.073). On the *bombina* side of the zone θ_b =0.0037, so selection on *variegata* marker alleles at this side of the zone is extremely weak at 0.001. However, the limits to θ_b are wide (0-0.06) and so a selection pressure of up to 0.023 is plausible. As stated above the limits of θ_v and θ_b overlap so they may be the same. The rates of decay here are similar to the Polish estimates; at Przemyśl θ_v = 0.017 and θ_b = 0.0094 and at Cracow θ_v = 0 (0-0.011) and θ_b = 0.011 (0-0.025); the limits at all transects overlap. However, because the effective selection across the transect described here is twice that in Poland, the selection acting directly on the marker enzymes is estimated to be greater; s_e = 0.0037 and 0.0016 in Przemyśl and Cracow, so the difference to Pešćenica in s_e is approximately ten-fold.

The stepped cline can be interpreted as a barrier to gene flow. The barrier may have a variety of sources. It could be due to linkage disequilibrium between the marker loci and those directly under selection, or it may be the result of some physical obstacle (Chapter 1). Whatever the source of the barrier its strength can be described by the ratio between the step in allele frequency across the centre and the gradient at the edge. The barrier to gene flow on the *bombina* side is large; it would take an equivalent neutral allele the same time to travel 191km (assuming a constant width of 4.67): approximately 34,500 generations ($T \approx [B/\sigma]^2$). However, the limits on this side of the zone are too wide (reflecting fewer samples) to say anything definite about the estimate here. Barrier strength on the *variegata* side of the zone is more reliable. Here B_v/w =3.3 (support limits, 1.46-8.07) so a neutral allele could require up to 225 generations to cross an equivalent barrier. The limits to θ and B on either side of the

zone overlap so that although the most likely estimates of barrier strength on either side of the zone are different (suggesting some asymmetry) the possibility that they are the same cannot be excluded. At Pešćenica the change in allele frequency across the step goes from 0.12 to 0.67 (a difference of 0.55), whereas in Poland the step is from 0.05 to 0.9 (difference of 0.85). The Polish transects show a greater change in gene frequency across the step (Fig. 6.1).

If the barrier to gene flow is due to selection on linked loci then $B = w(\overline{W}_{centre}/\overline{W}_{edge})^{1/r}$ (Barton 1986, Chapter 1). This can give an estimate of the net fitness of hybrids in the centre of the zone. The harmonic mean recombination rate between the marker loci and selected loci in *Bombina* has been estimated as 0.25 (Szymura and Barton 1991). As the limits to the barrier strength on the *bombina* side are so wide only the barrier strength on the *variegata* side will be used. $B_V/w = 3.3$ (1.5, 8.1). This means that the fitness of hybrids in the centre of the zone is estimated as 0.74. This is smaller than the Polish estimates where the change in gene frequency across the centre is greater. This indicates a stronger barrier to gene flow and hence a lower net fitness of hybrids at 0.58 and 0.65 for Przemyśl and Cracow respectively.

Differences between the Polish and the Pešćenica transects.

The estimates inferred above were based on the same assumptions that allowed inferences from the Polish cline to be made, i.e. a cline maintained by a simple balance between dispersal and selection.

The differences between the inferred and observed estimates between the present transect and those from Poland can be summarised as follows (values for the Pešćenica transect are given in Table 6.1, and for the Polish transects in Table 1.2):-

Differences in the observed patterns

1. Disequilibrium is approximately twice that estimated in Poland.
2. There is a significant deficit of heterozygotes in the centre of the zone, relative to Hardy-Weinberg proportions.
3. There is a habitat preference such that *bombina*-like individuals prefer ponds and *variegata* like individuals prefer puddles.
4. There is adaptation to habitat; *bombina* are relatively fitter in ponds than*variegata* (in terms of growth rate) and *variegata* are fitter than *bombina* in puddles.

5. There are differences in the shape of the cline.

   a. The step in gene frequency at the centre of the zone is shallower.
   b. The width of the cline varies from place to place.
   c. There is a significant difference in gene frequency between habitats across the cline.
   d. Unlike Poland there is not a smooth transition of genotypes across the zone. There is a noisier distribution of gene frequencies around the cline

Differences between inferred estimates

1. Dispersal is slightly greater (1.03km compared to 0.94km averaged across both transects in Poland).
2. The effective selection is approximately twice that in Poland.
3. Difference in selection acting on marker loci is approximately ten-fold (though the difference is not significant).
4. Difference in the estimated fitness of hybrids. This is reflected by the shallower step in gene frequency, which is interpreted as a weaker barrier to gene flow giving the estimated fitness of hybrids at Pešćenica as 74% (relative to pure tyes) compared to 62% averaged over both transects in Poland.

6.2 Accounting for the differences in the estimates of disequilibrium and heterozygote deficit and implications for dispersal estimates; the effect of a habitat preference and adaptation.

Disequilibrium

Disequilibrium is twice that seen in Poland. This also accounts for twice the estimated strength of the effective selection. Given that the estimated fitness of hybrids (inferred from the step in gene frequency) in Pešćenica is greater than hybrid fitness in Poland there is an apparent paradox. There are two possible sources of error.

1) The model used to calculate the effective selection implies selection is against heterozygotes. This may be inappropriate.

2) Disequilibrium is inflated above what it should be given that disequilibrium is maintained through a balance of dispersal and selection alone.

Given that adaptation to habitat does occur in this hybrid zone it is reasonable to argue that selection will not purely be operating against heterozygotes irrespective of their environment. In this case it might be more appropriate to estimate selection, for example, as Endler (1977) suggested, along an environmental gradient. However, both these models are dispersal-dependent and differences in the strength of selection estimated either way are negligible. As Moore and Price (1993) pointed out; biases in the strength of selection are far more likely to come from errors in dispersal estimates.

I believe the source of the error comes from the assumption that disequilibrium is generated solely by dispersal across the zone. Linkage disequilibrium will also be generated through a habitat preference. If the preference were perfect then each habitat would contain only one genotype and there would be no disequilibrium generated. The crucial point is that the preference is not perfect. There will also be mixing between habitats so some fraction of the population will be from the opposing genotype. This thesis provides clear evidence for a habitat preference (Chapter 4) and the wide variance of genotypes in the habitats demonstrates mixing between habitats. This will inflate disequilibrium over and above that generated via dispersal alone.

Disequilibrium is estimated from pairwise associations between neutral enzyme markers. In order to observe the preference the marker loci must be in disequilibrium with the loci controlling the preference. The preference may be due to imprinting and have no genetic basis (Chapter 5) but in this case one would expect associations between the observed preference and the marker loci to break down more quickly. Considering the length of time these taxa have been hybridising one would not expect such inflated levels of disequilibrium. Also, there is evidence that the preference is adaptive (Chapter 5). Each taxon is relatively fitter in its own habitat (within enclosures). Therefore it is highly likely that there is a strong genetic basis to the preference.

One way to test this would be to release laboratory reared animals into the field to see what preference (if any) was expressed. This was attempted with some juvenile *Bombina* reared in the laboratory but out of more than two hundred individuals released only one was recaught. Imprinting or learning of habitat type cannot be excluded and may reinforce any genetic basis. In general *variegata* populations were

found only in puddles on the *variegata* side of the zone while *bombina* were found predominantly in ponds on the *bombina* side (Chapter 3). This may reflect habitat availability or differential occupancy of equally available habitats. In either case the parental populations generally existed in one habitat type or the other so there was an opportunity for habitat conditioning.

Recombination with the opposing taxon would tend to break up the associations, whereas dispersal and selection (regardless of its nature) would maintain them. Given that the preference has a genetic component then linkage disequilibrium will be generated between the preference genes and those genes conferring fitness. As a result the marker genes will be in disequilibrium both with the genes under direct selection and the preference genes. However, computer simulations and a theoretical analysis would be needed to confirm this.

Deviations from Hardy-Weinberg proportions at Pešćenica

Another fundamental difference between populations across the Polish transects and those here is that populations within the hybrid zone at Pešćenica are not in Hardy-Weinberg proportions (only one population at Cracow showed a significant heterozygote deficit; this was attributed to three *bombina* genotypes in a predominantly *variegata* sample). The pattern of heterozygote deficit seen at Pešćenica is similar to that for disequilibrium; F_{is} is strong, reaching a maximum of 0.26 in the centre of the zone. It is significant in the centre only, unlike disequilibrium which is also significant at the edges. It is remarkable to see such a strong heterozygote deficit in the centre of the zone. This may come about in various ways:-

1. Unless the probability of choosing a particular habitat is very high it is likely that some individuals will choose the habitat of the opposing taxon. Therefore if a few *bombina*-like individuals disperse to a puddle where there is a predominantly *variegata*-like population, then a sample of adults will show a heterozygote deficit.

2. Individuals may mate at random within a population. There may be selection against offspring that are not adapted to the habitat. If these are predominantly heterozygotes and a large fraction of the offspring return to that site to breed, then sampling the adult population in the next generation will again show deviations from Hardy-Weinberg proportions.

3. Non-random mating within populations.

Although each of these explanations may contribute to the heterozygote deficit it is unlikely that the first two explanations will produce the observed estimate of F_{is} because the variance in gene frequency within hybrid populations is not large enough. Non-random mating is therefore likely. Direct evidence for this would come from a comparison of the gene frequencies of the adult population with that in the subsequent population of tadpoles. Just such a collection of adults and tadpoles was sampled from one site (1003) in 1992. Electrophoretic examination of the adult population showed that it is mainly *variegata*-like with a few very *bombina*-like individuals (Appendix 2.2 gives individual scores). A recent examination of allozymes in the tadpole population showed a similar pattern, (L.Kruuk; pers comm) i.e. they are mainly *variegata*-like with a few very-*bombina* like individuals. If there was random mating one would not expect any of the tadpoles to be as *bombina*-like as observed, because the probability of the few *bombina*-like individuals meeting at random must be low. The only explanation is therefore positive assortative mating.

One aspect of non-random mating is that in conjunction with the habitat preference it will effectively increase the linkage disequilibrium observed in any population. Also as there is positive assortment of genotypes between habitats due to the preference, mate choice may also act to reinforce the preference because choosing the right habitat will allow access to preferred mates.

It is theoretically possible that disequilibrium could be generated and maintained by the preference alone independent of dispersal. Diehl and Bush (1989) have shown that substantial disequilibrium can be generated between habitat preference genes and those conferring fitness in the habitat, when mating occurs within the habitat type. This is due to the fact that the disequilibria are generated between populations and are unaffected by recombination (1989). Here, disequilibrium is estimated within populations and may be generated by mixing between habitats which differ in gene freqency as a result of the habitat preference. Potentially, therefore, recombination will break down the associations within the habitats, though selection against hybrids and positive assortative mating will help maintain them.

It is probable that both dispersal and preference contribute to the associations. Not only is there direct evidence for dispersal (Chapter 4), but it is unlikely that such

strong disequilibrium would be seen as a result of habitat preference and selection in relation to habitat alone.

There will be a positive feedback between the sources of the disequilibrium. Selection helps generate diseqilibrium but will also reinforce both non-random mating and the preference, which will further increase disequilibrium and hence the effective selection on any locus will be increased even further and so on. Therefore it is likely that disequilibrium is generated and maintained not just through a dispersal selection balance but due to a balance of dispersal of parental genotypes into the zone, selection against hybrids, non-random mating and a preference between alternative habitats.

This means that disequilibrium in this hybrid zone will be inflated compared to those zones where no preference exists. This will therefore bias inferred estimates which rely on the assumption of dispersal and selection alone i.e. dispersal rate, the effective selection and the selection acting directly on each marker locus. Although the estimate of the net fitness of hybrids is deduced independently from the step in gene frequency at the centre of the zone, it is also biased as it is based on the interpretation that the barrier is generated through disequilibrium generated by dispersal alone. Allowing for a difference in gene frequency between habitats reduces the step. This would explain the paradox between the extremely strong estimate of the effective selection and yet the relatively high net fitness of hybrids.

The patterns of disequilibrium and heterozygote deficit show some asymmetry in relation to the change in gene frequency of different populations (Chapter 2). Both are stronger in *bombina*-like populations than they are in *variegata*-like ones. An asymmetric barrier to gene flow may be due to differences in fitness of the hybridising taxa or differences in population structure (such as density). In either case an asymmetric barrier implies that the zone is moving (Barton and Hewitt 1985). If there are differences in fitness one would expect the cline to move towards the less fit population. However a moving cline can be stabilised at an environmental transition. Where there is adaptation to different environments (as is the case here) then differences in habitat availability can stabilise the cline. If there are no ponds in the upland forest area then *bombina* may be effectively excluded. Assuming that the zone is stable then the asymmetry could arise via the following mechanisms:-

1. If selection against *variegata*-like individuals is greater on the *bombina* side of the zone then disequilibrium will be greater on the *bombina* side of the zone because

selection will counter recombination more on the *bombina* side than the *variegata* side. This would result in a steeper cline on the *bombina* side. There is evidence for this at Pešćenica (see below).

2. Differential dispersal.

Differences in dispersal should cause a tension zone to move until trapped by some barrier. If the barrier is a density trough then density and dispersal balance out and the zone is stable (e.g. Hewitt 1988, Fig. 3). In this situation one would not expect any difference in disequilibrium either side of the zone. However, if the zone is stabilised by an environmental barrier which does not correlate with a reduction in population density then there will be effectively more dispersal in one direction across the cline than the other. If *variegata* disperse at a greater rate than *bombina* then there will be a net flux of *variegata* like combinations of genes into the *bombina* side of the zone. Both this and the greater selection against *variegata*-like individuals on this side of the zone will result in a steeper rise in disequilibrium.

3. Habitat preference.

A habitat preference will inflate both disequilibrium and the observed heterozygote deficit when it is combined with mixing between habitat types. This effect will be strongest where there is mixture of habitat types. If habitat availability changes across the zone such that ponds are absent on the *variegata* side but both puddles and ponds are present on the *bombina* side then disequilibrium and F_{is} will be stronger on the *bombina* side of the zone.

A differential habitat preference would also result in asymmetric patterns of disequilibrium and F_{is}. If for example *variegata* shows a higher fidelity to puddles than *bombina* does to ponds then, within the hybrid zone, disequilibrium and the apparent F_{is} may be greater in puddles. This is because *bombina*-like individuals will be found in puddles but *variegata*-like individuals will not be found in ponds. Therefore different genotypes will be more often juxtaposed in puddles than in ponds. This argument does not rely on habitat availability because, as gene frequency increases, there will be fewer *bombina* like individuals available.

Implications for dispersal

The existence of a habitat preference was not demonstrated across the Polish transects. In contrast the habitat preference observed at Peščenica generates inflated disequilibrium over and above that expected for a dispersal dependent cline. This means that the estimate of dispersal inferred from the value of disequilibrium will also be inflated. However, a habitat preference will also affect dispersal more directly.

Dispersal rates can differ in two ways. Dispersal of either taxon may be impeded by a lack of suitable habitats or dispersal may differ between taxa independently of the habitat availability. Both these types of dispersal will have important effects on the shape of the cline. These are described below in Section 6.3.

If there were no habitat preference then dispersal rate would depend only on the differences between taxa and the habitat availability. However a habitat preference complicates the situation. Four types of dispersal can be identified.

1. Individuals may disperse between the same habitat types in the same area; for example *variegata* may move between neighbouring puddles.

2. Individuals may disperse between different habitats in the same area. Dispersal rates may differ depending on the starting position of the individual. Therefore the dispersal rate of *variegata* from a pond to a puddle may be greater than from a puddle to a pond.

3. Individuals may move to the same habitat type but in a different area. For example the dispersal rate of *variegata* from one puddle to another some distance away.

4. Individuals may move to a different habitat type in a different area. For example the dispersal rate of *variegata* from puddles to ponds some distance away or *vice versa* (see 2 above).

The rates of these different movements will also depend on habitat availability. For example, where there are more puddles than ponds the dispersal rate of *variegata* moving from a puddle to a pond may be less than when there are fewer puddles, or conversely, if there are fewer puddles in an area the dispersal rate of *variegata* from a pond may be less. Also, where there are fewer puddles in an area the total turnover of *variegata* individuals through that area may be greater than if puddles were abundant.

6.3 Accounting for differences in the shape of the cline

Differences in gene frequency between habitats

The model used to describe the distribution of genotypes across the zone is a stepped cline which allows for a difference in gene frequency according to habitat. The difference in gene frequencies between habitats is described by alpha (α) which increases in the same way in both habitat types and reaches a maximum in the centre. Here the difference between the two habitat types is $\alpha/2$. The most likely estimate of alpha is 0.31. This means that the difference in gene frequency between habitats at the centre of the zone is 0.15. The observed difference in gene frequency between habitats estimated by comparing nearby sites is 0.5 (Chapter 3). Explanations were provided in that chapter for the discrepancy. It was concluded that the true value of α was probably intermediate between the estimates.

Allowing for a difference in gene frequency between habitats significantly improves the likelihood of the model where habitat is not taken into account. This has important implications for inferences made from the size of the step in gene frequency in the centre of the zone (i.e. the barrier strength) and also for the width of the cline. These will be discussed below.

The step in gene frequency at the centre of the cline

The step in gene frequency at the centre of the cline is less in Pešćenica than in Poland (Fig. 6.1). The step in gene frequency can be interpreted as a barrier to gene flow where differences in the gradient of gene frequency at the edge and in the centre represent the barrier against an introgressing allele (Chapter 1).

Given this interpretation then estimates of barrier strength on the *bombina* and *variegata* sides of the zone at Pešćenica imply that the barrier to gene flow into *bombina* is greater than that into *variegata*. There is a tenfold difference between the most likely estimates of barrier strength either side of the zone. However the limits of barrier strength on the *bombina* side are so large that it is unreasonable to make any definite statement. There is some suggestion, however, that the estimates are asymmetric. This implies there is differential selection either side of the zone. This results in a cline which is steeper on the *bombina* side and may account for the asymmetric patterns of disequilibrium and heterozygote deficit mentioned above.

What is interesting is that the asymmetry implies that the net fitness of *bombina* hybrids on the lowland plains is relatively greater than *variegata*. This could provide a mechanism to explain the post-glacial expansion of *bombina* described by Arntzen (1978) at the expense of *variegata* as they spread across Bohemia, the Bohemian-Moravian plateaux and the Hungarian plains. However the barrier strength at the Polish transects provide conflicting evidence. There is little suggestion at Cracow for any asymmetry although the limits for both estimates are wide. At Przemyśl there is a similar pattern to that found at Pešćenica, the barrier to gene flow into *bombina* being less than into *variegata*. Unfortunately the estimates of gene flow into *bombina* are wide, so that again the results have to be interpreted with caution.

One has to be extremely careful about the interpretation of the 'barrier' to gene flow in the context of this transect. The barrier is a description of the shape of the cline in the centre of the zone. The net fitness of hybrids in the centre can be inferred from the size of the step given the width of the cline and the recombination rate between the marker and selected loci. Inferences from the step in gene frequency imply that the net fitness of hybrids is greater at Pešćenica than in the Polish transects. However this does not take into account habitat preference. Both the preference and adaptation to habitat imply that the two genotypes can co-exist in the centre of the zone. One could therefore envisage a very small step in gene frequency between populations in the centre ultimately leading to sympatry maintained by the habitat preference and assortative mating. Inferences from this 'barrier' strength would then imply that the fitness of hybrids was virtually the same as the pure types. However the populations would be effectively isolated from each other and no hybrids would be produced.

Variation in the width of the cline

There is significant variation in the width of the cline along its length (Chapter 3). Given the difference in habitats found between the two taxa the most parsimonious explanation for the variation in width is that it reflects the habitat availability of an area. Where there are few habitats sampled, combined with a close juxtaposition of the habitat types (for example at Dužica) then the cline becomes extremely narrow. Where many habitats are sampled and there is a wider distribution of habitat types, the transition between the two taxa and hence the width of the cline is greater. It is reasonable to assume that if one allows for a difference in gene frequency between

habitat types then some of the variation in width will be accounted for. However, even when habitat is included in the model, significant variation in width remains.

Differences in the availability of the two habitat types along the length of the cline are not the only exlanation for differences in width. A cline which is maintained in direct response to selection will be expected to vary in width and shape much more than one maintained in a dispersal selection balance (Karlin and Richter-Dyn 1976). However, width may vary substantially even within a dispersal dependent cline (Barton and Hewitt 1985). There are a number of reasons for this:

1. The width and shape of a cline may vary due to the inherent genetic variation within the taxa as they extend along the cline. This may create differences in the barrier to gene flow across different transects of the same hybrid zone. In particular, the differences between *variegata* subspecies may account for some of the differences between the shape of the cline at Pešćenica and those across the Polish transects. Although this might account for differences in the width of transects thousands of kilometres apart, the variation in width along the length of the cline examined at Pešćenica occurs over the much shorter distance of 36km. It is unlikely that there is enough genetic variation within taxa along this distance to account for the dramatic differences in width.

2. Hybrid zones may be 'modified' in different ways in different regions. For example if hybrid zones are areas which promote reproductive isolation then it could be envisaged that the rate at which isolation occurs differs due to differences in local environmental conditions or mutation rates. This will alter the barrier to gene flow between the taxa. If the fitness of hybrids changes sufficiently between areas then it will be perceived through variation in width between transects. Again, given the short length of the cline at Pešćenica, this is unlikely to account for differences in width.

3. In any zone maintained by a dispersal-selection balance changes in density and dispersal are likely to alter the width and shape of the cline. This is the most plausible explanation for the variation in width seen along the cline at Pešćenica.

If habitat availability differs along the length of the cline then dispersal rates and the density of populations will differ between areas. It is conceivable that along the 36km describing the length of the cline the puddle and pond distribution will differ. Where there is a deficit of both habitat types the population density in that area will be reduced

and the cline will narrow. This may occur at Dužica. Where there is a higher frequency of habitats the local population density will be effectively increased and the cline will widen, for example at Lekenik.

The position of the cline will also be affected by habitat availability. The course of the cline follows gross differences in habitat. In particular it bulges around the area of lowland forest to the north of Pešćenica. Since there is selection in relation to the habitat type, habitat availability will directly affect the course of the cline. Particular genotypes will predominate in areas with a high proportion of one habitat type. The bulge in the cline represents an area of lowland forest where populations were predominantly sampled from puddles. This *variegata*-like habitat would explain why the zone bows towards the *bombina* side of the zone

Selection may also create a 'hybrid sink' (Barton 1980, Barton and Hewitt 1985). In effect this causes a reduction in density due to the elimination of individuals in the centre of the zone. Where there is extrinsic selection, by environmental factors, the maladapted genotypes will be removed. Habitat availability will therefore have an effect on the relative fitness of genotypes within an area. In this study habitats have been classified into either puddles or ponds. No doubt this is an oversimplification as it is unreasonable to expect a complex heterogeneous environment to have only two habitat types. Puddles and ponds may be at the extreme ends of a scale which may display a whole range of intermediate habitat types. One could speculate that if intermediate habitat types are available hybrid individuals may not be selected against as strongly as in the extreme types. The fitness of hybrids, which may vary from place to place depending on the abundance of these intermediate habitats, will affect population density.

Explaining the residual variation

Allowing for both habitat and variation in width significantly improves the likelihood of the model describing the cline. However there is significant residual variation around the cline even when these factors are taken into account. Allowing for fluctuations between loci reduces the residual variation but there is still a large and significant scatter around the cline (Fig. 6.1). There is not the smooth transition of genotypes across the zone as seen in Poland.

Part of the excess variation may be due to constraining α to be the same across the zone. It may be that the difference in the gene frequencies of populations in different habitats is greater on one side of the zone than the other. This could be due to a differential habitat preference or due to a different availability of habitats across the zone.

Allowing for variation in width may account for some differences in habitat availability or relative population densities but there is a constraint on how much the cline can vary when it is described by a set number of segments. In Chapter 2 it was shown that twelve segments significantly improved the fit. However, more often than not the cline then curled unrealistically. When so many sites and so many parameters are involved it will always be difficult to find the optimum solution.

6.4 Adaptation to habitat and the nature of selection

Some of the inferences made from the Cracow and Przemyśl transects were made under the assumption that selection was acting against recombinants irrespective of the environment. This assumption produced plausible values for the various parameters estimated. Over and above this Szymura and Barton (1986) had direct evidence that selection was acting against hybrids rather than in relation to the environment. For example hybrid individuals from the zone at Cracow showed increased developmental abnormalities not present in the pure populations and tadpoles collected in the field showed various deformations of the tooth row pattern thought to impair feeding. Other researchers working in the area had produced similar results (reviewed in Szymura and Barton 1986, Szymura 1993). This, combined with the fact that there was only a broad association with the environmental transition, provided evidence that adaptation was not directly maintaining the cline.

Clear evidence is provided in this thesis to show that there is selection in relation to the environment (Chapter 5). In addition, unlike the Polish transects, there is no direct evidence for selection against hybrids. Laboratory crosses within hybrid and pure populations collected from the Pešćenica transect showed that hybrids were as viable as the parental populations (Nürnberger *et al.* 1993). There were no differences in survival prior to or post metamorphosis. Interestingly, offspring from some crosses between pure types, i.e. F1 individuals did show a marked reduction in viability. Their contribution to the maintenance of the hybrid zone however is probably minimal; the presence of F1 individuals is rare in nature. Not one individual that was scored for

all the diagnostic enzymes at this transect could be interpreted as an F1 (Chapter 2). Most recombinants are the result of numerous backcrosses. Also, it may be that incompatible combinations of alleles affecting viability would be strongly selected against in the first generation and therefore not transmitted to future hybrid generations (Nürnberger *et al.* 1993). However one cannot discount selection against hybrids on the results of these laboratory crosses alone. It is likely that selection will act against hybrids both in relation to external environmental conditions and internal genetic incompatibilities. More field experiments are required to look at hybrid viability under a range of different environmental conditions.

Explaining the differences between the Polish and Pešćenica transects

One obvious question is why the Pešćenica transect is so different to those described in Poland, where there is neither a close association between genotype and any ecological variable, nor a habitat preference. There are three possible explanations:-

1. The taxa at Pešćenica are genetically differentiated from those in Poland (Szymura 1993). The Polish transects described a hybrid zone between the northern form of *B. bombina* and the Carpathian form of *variegata* while the hybrid zone at Pešćenica is between the southern form of *B. bombina* and the western form of *variegata*. As these populations of *variegata* may have diverged in allopatry (Chapter 1) there may well be fundamental differences between them in hybrid viability, adaptation and habitat preference.

2. The habitat distribution is different. In Poland it may be that the habitat availability is the same in each parental environment. If so the parental populations may make use of both habitat types, in which case imprinting may not occur and the preference expressed in the hybrid zone may be smaller or absent. At the other extreme, the transition of habitats may be so sharp that the majority of hybrid populations do not have the opportunity to select different habitat types. In this case selection against hybrids will be the predominant force determining the distribution of genotypes rather than selection and preference. A third possibility is that there may be a smooth transition of one habitat type to the other across the zone .

3. There may be different selection pressures acting on the hybrid zones in Poland and Croatia. Even if there is a habitat preference in Poland it may not confer the same

fitness advantage that it does in Croatia. One reason for this is that the transects in Croatia are in a warmer drier climate. Lowland puddles in Croatia may be more ephemeral than in Poland, imposing more constraints on development time within the hybrid zone. Nürnberger *et al.* (1994) have shown that the cline in development time and for egg size at Pešćenica is offset from the clines for the other quantitative traits. This may be due to direct and strong selection on that trait. The cline for development time shows that the minimum development times are for *variegata*-like hybrids rather that pure *variegata* . It may be that selection pressure for faster development rate is actually increased within the hybrid zone. One might even extrapolate from this and propose that *variegata*-like hybrids are sometimes favoured in parts of the hybrid zone as in the manner proposed by Moore (1977).

6.5 Consequences for speciation

A difference in fitness between habitats and an active habitat preference promotes assortative mating, as individuals of similar genotype will tend to be found in the same habitat. Indeed it has been shown in laboratory populations of *Drosophila melanogaster* that artificial disruptive selection on a habitat preference with genetically determined differences in fitness between habitats can bring about almost complete reproductive isolation between two subpopulations (Rice and Salt, 1988). The essential component of this experiment is that dispersal to the preferred habitat occurs prior to mating. This is important; many models of sympatric speciation do not take into account the association between habitat preference and mating site (Tauber and Tauber, 1989) and therefore do not allow for the positive assortative mating due to the habitat differences. Also, many models show that divergence between populations would only occur if there were strong physical linkage between the preference and fitness genes (reviewed by Tauber and Tauber 1989). Recent models have been biologically more realistic (allowing dispersal prior to mating) and have not invoked a physical linkage between preference and fitness loci. Bush and Diehl's model (1989) shows that linkage disequilibrium between preference genes and those conferring adaptation will also promote divergence. The *Bombina* hybrid zone at Pešćenica reflects many of these important issues. Here dispersal to a preferred habitat occurs prior to mating and there is a genetically based fitness difference between these habitats. The presence of strong linkage disequilibrium within the zone reflects the effects of selection and preference on these populations. There will be positive

assortative mating between habitat types due to the preference but this will also be enhanced by the non-random mating seen within habitat types.

Although there is a great deal of similarity between the *Bombina* situation at Pešćenica and the host-race models, they are different in many respects. Linkage disequilibrium is estimated within populations rather than between habitats. It is inflated by the habitat preference because some individuals go to the wrong habitat type. Likewise the heterozygote deficit may be due not to assortative mating between habitats, but to non-random mating within habitats. Linkage disequilibrium is inflated by the preference and non-random mating but much of it will also be generated and maintained by dispersal of the parental genotypes into the zone.

Although the differences between the Polish transects and the Pešćenica transect may have been generated when populations were isolated in allopatry, it is unknown whether the adaptation to habitat and the preference displayed across the Pešćenica transect arose during a period of geographic isolation, or occurred when the populations were in secondary contact. Many of the models which allow sympatric speciation to occur invoke an adaptive habitat difference. Disruptive selection causes divergence of the populations involved. This process will occur at a much faster rate if there is a habitat preference (Maynard Smith 1966, Bush and Diehl 1989). Climatic differences between southern and northern Europe may have promoted the occurrence of the preference at Pešćenica. It could be envisaged that the same process is happening in Poland but at a slower rate.

If the differences among the Polish transects and across the hybrid zone at Pešćenica each occurred in parapatry then the hybrid zone described here may be evidence for the reinforcement of reproductive isolation. Reinforcement is the process by which prezygotic barriers to gene exchange are improved by natural selection (Howard 1993). It occurs in the presence of gene flow between two populations (Butlin 1989). The habitat preference displayed by the two hybridising taxa of *Bombina* at Pešćenica could be considered such a barrier. In this sense the inflated estimate of the effective selection acting to maintain the barrier to gene flow between *bombina* may be a true reflection of 'the reproductive isolation' between the taxa. Although the net fitness of hybrids at Pešćenica is greater than that estimated for the Polish transects one could envisage the two taxa living in sympatry where there is separation by habitat and no gene flow as a result.

Other hybrid zones, especially mosaic ones, have demonstrated a close correlation between genotype and habitat (e.g. Harrison and Rand 1989, Sites *et al.* in press; see Chapter 1). The distribution of genotypes in these zones implies that there is extrinsic selection, where the barrier to gene flow is due to selection in relation to the environment as well as selection against hybrids. However, neither has provided evidence for a habitat preference. The hybrid zone between chromosomal races of the lizard *Scleropus* (Sites *et al.*, in press) excluded the possibility of a preference as there was no small scale association between habitat and genotype. This allowed robust inferences to be made from parameters such as linkage disequilibrium. Harrison and Rand (1989) emphasised the importance of a patchy habitat on the distribution of genotypes in the cricket hybrid zone they describe. They suggest that mosaic hybrid zones would be a more favourable site for reinforcement than narrow clines where there is a smooth transition between genotypes. However, a habitat preference may provide the mechanism for re-inforcement in a more traditional gradient model. The hybrid zone at Peščenica can still be described by a cline although the distribution of genotypes within it is patchy. It would be interesting to know what effect a habitat preference would have in a mosaic hybrid zone and what sort of levels of disequilibrium would be generated.

It is apparent from this thesis that a hybrid zone may act as a barrier to gene flow in many different ways. It is important to be aware that barriers to gene flow between populations may have genetic, ecological and behavioural components. The hybrid zone at Peščenica highlights the importance of environmental heterogeneity in determining and maintaining its position and genetic structure. Perhaps the most revealing part of this analysis comes from the comparison of this hybrid zone with those in Poland. The key to the differences was identified via the inflated disequilibrium. The hybrid zone at Peščenica is maintained through a balance of dispersal and selection, but with a barrier to gene flow that is strengthened by habitat preference. Both selection and a habitat preference will contribute to the reproductive barrier between the taxa. The habitat preference analysed here throws new light on the interpretation of clines.

REFERENCES

Abernethy, K. (1994) The introduction of Sika deer *(Cervus nippon nippon)* to Scotland. Ph.D thesis, University of Edinburgh.

Aguade, M., Miyashita, N.& Langley, C.H. (1992) Polymorphism and divergence in the Mst26A male accessory gland gene region in *Drosophila*. *Genetics* 132: 755-770.

Arntzen, J.W. (1978) Some hypotheses on postglacial migrations of the firebellied toad, *Bombina bombina* (L) and the yellowbellied toad, *Bombina variegata* (L). *J.Biogeogr.* 5: 339-395.

Baker, R.J. (1981) Chromosome flow between chromosomally characterized taxa of a volant mammal, *Uroderma bilobatum* (Chiroptera: Phyllostomatidae). *Evolution* 35: 296-305.

Barton, N.H. (1979a) The dynamics of hybrid zones. *Heredity* 43: 341-359.

Barton, N.H. (1979b) Gene flow past a cline. *Heredity* 43: 333-339.

Barton, N.H. (1982) The structure of the hybrid zone in *Uroderma bilobatum* (Chiroptera: Phyllostomatidae). *Evolution* 36: 863-866.

Barton, N.H. (1983) Multilocus clines. *Evolution* 37: 454-471.

Barton, N.H. (1986a) The effects of linkage and density-dependent regulation on gene flow. *Heredity* 57: 415-426.

Barton, N.H. (1986b) The maintenance of polygenic variation through a balance between mutation and stabilising selection. *Genet.Res.* 47: 209-216.

Barton, N.H. (1988) Speciation. In A. Myers & P. Giller (eds.), *Analytical Biogeography*: 185-218. London: Chapman and Hall.

Barton, N.H.& Bengtsson, B.O. (1986) The barrier to genetic exchange between hybridising populations. *Heredity* 57: 357-376.

Barton, N.H.& Gale, K.S. (1993) Genetic analysis of hybrid zones. In R.G. Harrison (ed.), *Hybrid zones and the Evolutionary Process*: 13-45. New York: Oxford University Press.

Barton, N.H.& Hewitt, G.M. (1981a) A chromosomal cline in the grasshopper *Podisma pedestris*. *Evolution* 35: 1008-1018.

Barton, N.H.& Hewitt, G.M. (1981b) The genetic basis of hybrid inviability between two chromosomal races of the grasshopper *Podisma pedestris*. *Heredity* 47: 367-383.

Barton, N.H.& Hewitt, G.M. (1981c) Hybrid zones and speciation. In W.R. Atchley & D. Woodruff (eds), *Evolution and Speciation: Essays in honour of M.J.D White*: 109-145. Cambridge: Cambridge University Press.

Barton, N.H.& Hewitt, G.M. (1985) Analysis of hybrid zones. *Ann.Rev.Ecol. & Syst.:* 16: 113-148.

Barton, N.H.& Hewitt, G.M. (1989) Adaptation, speciation and hybrid zones. *Nature* 341: 497-503.

Bazykin, A.D. (1969) Hypothetical mechanism of speciation. *Evolution* 23: 685-687.

Begon, M. (1979) In *Investigating animal abundance.* London: Edward Arnold.

Berven, K.A. (1981) The genetic basis of altitudinal variation in the wood frog *Rana sylvatica.* I. An experimental analysis of life history traits. *Evolution* 36: 962-983.

Blackwell, J.M.& Bull, C.M. (1978) A narrow hybrid zone between the Western Australian frog species, *Ranidella insignifera* and *R. pseudo-insignifera:* the extent of introgression. *Heredity* 40:13-25

Blaxter, J.H.S.& Hempel, G. (1963) The influence of egg size on herring larvae (*Clupea harengus* L.). *J. de Conseil.* 28: 211-240.

Boulenger, G.A. (1886) On two European species of Bombinator. *Proc. Zool. Soc. Lond.* 1886: 499-501.

Butlin, R.K. (1989) Reinforcement of premating isolation. In D.Otte & J. Endler (eds.), *Speciation and its consequences*: 158-179. Sunderland, Massachusetts: Sinauer Associates, Inc.

Butlin, R.K.& Hewitt, G.M. (1985) A hybrid zone between *Chorthippus parallelus parallelus* and *C p. erythropus* (Orthoptera: Acrididae): behavioural characters. *Biol. J. Linn. Soc.* 26: 269-285.

Carson, H.L. (1985) Unification of speciation theory in plants and animals. *Syst.Bot.* 10: 380-390.

Carson, H.L.& Kaneshiro, K.Y. (1976) *Drosophila* of Hawaii: systematics and ecological genetics. *Ann.Rev.Ecol. Syst.* 7: 311-346.

Carson, H.L., Nair, P.S.& Sene, F.M. (1975) *Drosophila* hybrids in nature: proof of gene exchange between sympatric species. *Science* 189: 806-807.

Charlesworth, B., Coyne, J.A.& Barton, N.H. (1987) The relative rates of evolution of sex chromosomes and autosomes. *Amer.Nat.* 129: 113-146.

Clarke, B.C. (1966) The evolution of morph ratio clines. *Amer. Nat.* 100: 389-400.

Coyne, J.A. (1985) The genetic basis of Haldane's Rule. *Nature* 314: 736-738.

Coyne, J.A., Orr, H.A.& Futuyma, D.J. (1988) Do we need a new species concept ? *Syst. Zool.* 37: 190-200.

Coyne, J.A. (1992) Genetics and speciation. *Nature* 355: 511-515.

Coyne, J.A.& Orr, H.A. (1989) Two rules of speciation. In D.Otte & J. Endler (eds.), *Speciation and its consequences*: 180-207. Sunderland, Massachusetts: Sinauer Associates, Inc.

Cracraft, J. (1989) Speciation and its Ontology. In D.Otte & J. Endler (eds.), *Speciation and its consequences*: 28-59. Sunderland, Massachusetts: Sinauer Associates, Inc.

Czopkowa, G.& Czopek, J. (1955) The vascularisation of the respiratory surfaces in *Bombina variegata*. *Bulletin Academia Polonica Scientatis* Cl II: 313-318.

Diehl, S.R.& Bush, G.L. (1989) The role of habitat preference in adaptation and speciation. In D.Otte & J. Endler (eds.), *Speciation and its consequences*: 345-365. Sunderland, Massachusetts: Sinauer Associates, Inc.

Dillon, W.R.& Goldstein, M. (1984) In *Multivariate Analysis - Methods and Applications*. New York: John Wiley and Sons.

Dobzhansky, T. (1940) Speciation as a stage in evolutionary divergence. *Amer. Nat.* 74: 312-321.

Dobzhansky, T. (1970) *Genetics of the Evolutionary Process*. New York: Columbia University Press.

Duellman, W.E.& Trueb, L. (1986) *Biology of Amphibians*. Baltimore and London: The Johns Hopkins University Press.

Edwards, A.W.F. (1972) *Likelihood*. Cambridge: Cambridge University Press.

Endler, J.A. (1977) *Geographic Variation, Speciation, and Clines*. Princeton: Princeton University Press.

Endler, J.A. (1980) Natural selection on colour patterns in *Poecilia reticulata*. *Evolution* 34: 76-91.

Endler, J.A. (1983) Natural and sexual selection on color patterns in poecilid fishes. *Environmental biology of fishes* 9: 173-190.

Fisher, R.A. (1925) *Statistical methods for Research Workers*. Edinburgh: Oliver and Boyd.

Fisher, R.A. (1937) The wave of advance of advantageous genes. *Ann. Eugenics* 7: 355-369.

Futuyma, D.J.& Mayer, G.C. (1980) Non-allopatric speciation in animals. *Syst.Zool.* 29: 254-271.

Gartside, D.F. (1980) Analysis of a hybrid zone between chorus frogs of the *Pseudacris nigrita* complex in the southern US. *Copeia* 1980: 56-66.

Gauch, H.G. (1982) *Multivariate Analysis in Community Ecology*. Cambridge: Cambridge University Press.

Golding, G.B. (1984) The sampling distribution of linkage disequilibrium. *Genetics* 108: 257-274.

Gollmann, G. (1984) Allozymic and morphological variation in the hybrid zone between *Bombina bombina* and B *variegata* (Anura: Discoglossidae) in north-eastern Austria. *Z.Zool.Syst.Evol.Forsch.* 22: 51-64.

Gollmann, G. (1986) Genetic analysis of Bombina hybrids from eastern Slovakia. In Rocek. Z (ed.), *Studies in Herpetology:* 121-126. Prague: Charles University Press.

Gosner, K.L. (1960) A simplified table for staging anuran embryos and larvae with notes on identification. *Herpetologica* 16: 183-190.

Grant, P.R.& Grant, B.R. (1989) Sympatric speciation and Darwin's Finches. In D. Otte & J. Endler (eds.), *Speciation and its consequences*: 433-457. Sunderland, Massachusetts: Sinauer Associates, Inc.

Grant, V. (1971) *Plant Speciation*. New York: Columbia UniversityPress.

Greenwood, P.H. (1974) Cichlid fishes of Lake Victoria, East Africa: the biology and evolution of a species flock. London: British Museum (Natural History)

Gyllensten, U.& Wilson, A.C. (1987) Interspecific mitochondrial DNA transfer and the colonization of Scandinavia by mice. *Genet. Res.* 49: 25-29.

Hacking, I. (1965) *Logic of statistical inference*. Cambridge: Cambridge University Press.

Haldane, J.B.S. (1922) Sex ratio and unisexual sterility in hybrid animals. *J. Genetics* 12: 101-109

Haldane, J.B.S. (1948) The theory of a cline. *J. Genetics* 48: 277-284.

Harrison, R.G. (1990) Hybrid zones: windows on evolutionary process. *Oxford Surveys in Evolutionary Biology* 7: 69-128.

Harrison, R.G. (1993) Hybrids and hybrid zones: historical Perspective. In R.G. Harrison (ed.), *Hybrid zones and the Evolutionary Process*: 3-11. New York: Oxford University Press.

Harrison, R.G.& Rand, D.M. (1989) Mosaic hybrid zones and the nature of species boundaries. In D. Otte & J. Endler (eds.), *Speciation and its consequences*:. Sunderland, Massachusetts: Sinauer Associates, Inc.

Hedrick, P.W. (1987) Gametic disequilibrium measures: proceed with caution. *Genetics* 117: 331-341.

Hewitt, G.M. (1985) The structure and maintenance of hybrid zones with some lessons to be learned from alpine grasshoppers. In J. Gosalves, C. Lopez-

fernandez & C.G.de la Vega (eds),*Orthoptera: 15-54*. Madrid: Fundacion Ramon Areces.

Hewitt, G.M. (1988) Hybrid zones: natural laboratories for evolutionary studies. *Trends Ecol.Evol.* 3: 158-167.

Hewitt, G.M. (1989) The subdivision of species by hybrid zones. In D.Otte & J. Endler (eds.), *Speciation and its consequences*: 85-110. Sunderland, Massachusetts: Sinauer Associates, Inc.

Hewitt, G.M.& Barton, N.H. (1980) The structure and maintenance of hybrid zones as exemplified by *Podisma pedestris*. In R. L. Blackman, G. M. Hewitt, M.A. Ashburner (eds), *Insect Cytogenetics*. London: Symp. R. Entomol. Soc. 10:149-70.

Hewitt, G.M., Gosalvez, J., Lopez-Fernandes, C., Ritchie, M.G., Nichols, W.& Butlin, R.K. (1988) Differences in the nucleolar organisers, sex chromosomes and Haldane's Rule in a hybrid zone. In P.E. Brandham (ed.), *Kew Chromosome Conference III*. London: Her Majesty's Stationary Office.

Heyer, W.R., Donnelly, M.A., McDairmid, R.W., Hayek, L.-A.C.& Foster, M.S. (1994) *Standard Methods for Amphibians*. Washington and London: Smithsonian Institution Press.

Hill, W.G. (1974) Estimation of linkage disequilibrium in randomly mating populations. *Heredity* 33: 229-239.

Hill, W.G. (1975) Linkage disequilibrium among multiple neutral alleles produced by mutation in a finite population. *Theor.Pop.Biol.* 8: 117-126.

Hill, W.G.& Robertson, A. (1968) Linkage disequilibrium in finite populations. *Theor.Appl.Genet.* 38: 226-231.

Horbulewicz, L. (1927) Die geographisch Verbreitung der Bombinator- und Triton-Arten im Bereiche der Bezirke Sambor, Drohobycz, Stryj (Kleinpolen). *Bull. Int. Acad. Polon. Sci. Lett. Cracovie* B: 87-112.

Houde, A.E. (1988) Genetic differentiation in female choice between two guppy populations. *Anim. Behav* 36: 511-16.

Howard, D.J. (1986) A zone of overlap and hybridisation between two ground cricket species. *Evolution* 40: 34-43.

Howard, D.J. (1993) Reinforcement: Origin, Dynamics, and Fate of an Evolutionary Hypothesis. In R.G. Harrison (ed.), *Hybrid zones and the Evolutionary Process*: 13-45. New York: Oxford University Press.

Jackson, K.S. (1992) The population dynamics of a hybrid zone in the alpine grasshopper *Podisma pedestris*: an ecological and genetical investigation. Ph.D thesis, University College London.

Jaenike, J. (1988) Effects of early adult experience on host selection in insects: some experimental and theoretical results. *J. Insect Behav.* 1: 3-15.

Jaenike, J.& Holt, R.D. (1991) Genetic variation for habitat preference: evidence and explanations. *Amer.Nat* 137, Supplej: S67-S90.

Johnson, M.S. (1982) Polymorphism for direction of coil in *Partula suturalis*: behavioural isolation and positive frequency dependent selection. *Heredity* 49: 145-151.

Johnson, M.S., Murray, J.& Clarke, B. (1993) The ecological genetics and adaptive radiation of *Partula* on Moorea. *Oxford Surveys in Evol.* 167-237.

Jolly, G.M. (1965) Explicit estimates from capture-recapture data with both death and immigration - stochastic model. *Biometrika* 52: 225-247.

Kaplan, R.H. (1980a) Ontogenetic energetics in *Ambystoma. Physiol. Ecol.* 53: 43-56.

Kaplan, R.H. (1980b) The implications of ovum size variability for offspring fitness and clutch size within several populations of salamanders (*Ambystoma). Evolution* 34: 51-64.

Kaplan, R.H. (1987) Developmental plasticity and maternal effects of reproductive characters in the frog, *Bombina orientalis. Oecologia* 71: 273-279.

Key, K.H.L. (1968) The concept of stasipatric speciation. *Syst.Zool.* 17: 14-22.

Kirkpatrick, S., Gelatt, C.D.& Vecchi, M.P. (1983) Optimization by simulated annealing. *Science* 220: 671-680.

Laan, R.& Verboom, B. (1990) Effects of pool size and isolation on amphibian communities. *Biol.Conser.* 54: 251-262.

Lac, J. (1961) Verbreitung der Unken (Tiefland-Unke *Bombina bombina* L. und Berg-Unke *B, variegata* L.) in der Slowakei und Problematik deren gegenseitigen Kreuzung [in Slovak with German summary]. *Bio. Prace SAV Bratislava* 7: 5-32.

Lewontin, R.C. (1988) On measures of gametic disequilibrium. *Genetics* 120: 849-852.

Lorcher, K. (1969) Vergleichende bio-akustishe Untersuchungen an der Rot-und Gelbbauchunke, *Bombina bombina* (L.) and *B.v.variegata* (L.). *Oecologia.* 3: 84-124.

Madej, J. (1965) Variations in the sacral region of the spona in *Bombina bombina*, (L. 1761) and *Bombina variegata* (L. 1758) (Salientia, Discoglossidae). *Acta Bio.(Cracow) Ser. Zool. Cracow* 8: 185-197.

Madej, Z. (1964) Studies on the fire-bellied toad (*Bombina bombina*, L. 1761) and yellow-bellied toad (L. 1758) of Upper Silesia and Moravian Gate. *Acta Zoologica Cracow* 9: 291-336.

Mallet, J.L.B. & Barton, N.H. (1989a) Inference from clines stabilized by frequency-dependent selection. *Genetics* 122: 967-976

Mallet, J.L.B. & Barton, N.H. (1989b) Strong natural selection in a warning color hybrid zone. *Evolution* 43: 421-431.

Mallet, J.L.B., Barton, N., Lamas, G.M., Santisteban, J.C., Muedas, M.M. & Eeley, H. (1990) Estimates of selection and gene flow from measures of cline width and linkage disequilibrium in *Heliconius* hybrid zones. *Genetics* 124: 921-936.

Maxson, L.R.& Szymura, J.M. (1984) Relationships among discoglossid frogs: an albumin perspective. *Amphibia Reptilia* 5: 245-252.

May, R.M., Endler, J.A.& McMurtrie, R.E. (1975) Gene frequency clines in the presence of selection opposed by gene flow. *Amer. Nat.* 109: 659-676.

Mayr, E. (1942) *Systematics and the Origin of Species*. NewYork: Columbia University Press.

Mayr, E. (1957) *The species problem.* In *A symposim Presented at the Atlanta Meeting of the American Association for the Advancement of Science, December: 28-29, 1955*. American Association for the Advancement of Science. Washington DC.

Mayr, E. (1963) *Animal Species and Evolution*. Cambridge, Mass. Harvard University Press.

Mertens, R. (1928) Zur Naturgeschichte der Europäischen Unken *(Bombina). Z.Morphol.Ökol. Tiere* 11: 613-623.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A.& Teller, E. (1953) Equation of state calculations by fast computing machines. *J. Chem. Phys* 21: 1087.

Michalowski, J. (1961) Studies on species characteristics in *Bombina variegata* (L.) and *Bombina bombina* (L.). Applying the L:T indicator to the classifying purposes. *Acta Zool. Cracow* 51-59.

Moore, J.A. (1957) An embryologists view of the species problem. In E. Mayr (ed), *The species problem.* Amer. Assoc. Advancement Sci. Washington.

Moore, W.S. (1977) An evaluation of narrow hybrid zones in vertebrates. *Q.Rev.Biol.* 52: 263-278.

Moore, W.S.& Price, J.T. (1993) Nature of selection in the northern flicker hybrid zone and its implications for speciation theory. In R.G. Harrison (ed.), *Hybrid*

zones and the Evolutionary Process: 196-225. New York: Oxford University Press.

Moran, C., Wilkinson, P.& Shaw, D.D. (1980) Allozyme variation across a narrow hybrid zone in the grasshopper *Caledia captiva*. *Heredity* 44: 69-81.

Muller, H.J. (1942) Isolating mechanisms, evolution and temperature. *Biol.Symp* 6: 71-125.

Murray, J.& Clarke, B.C. (1980) The genus *Partula*: speciation in progress. *Proc.Roy.Soc.Lond. B* 211: 83-117.

Nagylaki, T. (1975) Conditions for the existence of clines. *Genetics* 80: 595-615.

Nagylaki, T. (1976) Clines with variable migration. *Genetics* 83: 867-886.

Nei, M. (1972) Genetic distance between populations. *Amer.Nat.*106: 283-292.

Nevo, E. (1973) Adaptive colour polymorphism in cricket frogs. *Evolution* 27: 353-367.

Nichols, R.A. (1985) Genetic and ecological differentiation across a hybrid zone. in J. Gosalves, C. Lopez-Fermnandes, and C. Garcia de la Vega (eds), *Orthoptera* Fundacion Ramon Areces, Madrid.

Nichols, R.A.& Hewitt, G.M. (1986) Population structure and the shape of a chromosomal cline between two races of *Podisma pedestris* (Orthoptera: Acrididae). *Biol.J.Linn.Soc.* 29: 301-316.

Nichols, R.A.& Hewitt, G.M. (1988) Genetical and ecological differentiation across a hybrid zone. *Ecol. Entomol.* 13: 39-49.

Nürnberger, B., Barton, N., MacCallum, C., Gilchrist, J.& Appleby, M. (in press) Natural selection on quantitative traits in the *Bombina* hybrid zone. *Evolution.*

Orr, H.A. (1992) Mapping and characterization of a 'speciation' gene in *Drosophila. Genet. Res.* 59: 73-80.

Orr, H.A. (1993) Haldane's Rule has multiple genetic causes. *Nature* 361: 532-533.

Orr, H.A. (1994) A mathematical model of Haldane's Rule. *Evolution* 47: 1606-1611.

Partridge, L. (1978) Habitat selection. In J.R. Krebs & N.B. Davies (eds), *Behavioural Ecology*: 351-376. Oxford: Blackwell Scientific Publications.

Paterson, H.E.H. (1978) More evidence against speciation by reinforcement. *S.Afr.J.Sci.* 74: 369-391.

Paterson, H.E.H. (1982) Perspective on speciation by reinforcement. *S.Afr.J.Sci.* 78: 53-57.

Paterson, H.E.H. (1985) The recognition concept of species. In S. Vrba (ed.), *Species and Speciation.* Transvaal Mus. Monogr. 4: 21-29.

Pollack, K.H., Nichols, J.D., Hines, J.E.& Brownie, C. (1990) Statistical inference for capture-recapture experiments. *Wildlife Monogr.* 107

Price, T., Turelli, M.& Slatkin, M. (1993) Peak shifts produced by correlated response to selection. *Evolution* 47: 280-290.

Provine, W. (1986) *Sewall Wright and Evolutionary Biology.* Chicago: University of Chicago Press.

Pulliam, H.R.& Danielson, B.J. (1991) Sources, sinks, and habitat selection; a landscape perspective on population dynamics. *Amer. Nat.* 137, Suppl. S50-S66.

Rafinska, A. (1991) Reproductive biology of the fire-bellied toads, *Bombina bombina* and *B. variegata* (Anura: Discoglossidae) : egg size, clutch size and larval period length differences. *Biol. J. Linn. Soc.* 43: 197-210.

Rand, D.M.& Harrison, R.G. (1989) Ecological genetics of a mosaic hybrid zone: mitochondrial, nuclear, and reproductive differentiation of crickets by soil type. *Evolution* 43: 432-449.

Rausher, M.D. (1984) The evolution of habitat preference in subdivided populations. *Evolution* 38: 596-608.

Rice, W.R. (1987) Speciation via habitat specialization: the evolution of reproductive isolation as a correlated character. *Evol. Ecol.* 1: 301-314.

Rice, W.R.& Salt, G.W. (1988) Speciation via disruptive selection on habitat preference: experimental evidence. *Amer. Nat* 131: 911-917.

Sanderson, N., Szymura, J.M.& Barton, N.H. (1992) Variation in mating call across the hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata. Evolution* 46: 595-607.

Schliewen, U.K., Tautz, D.& Pääbo, S. (1994) Sympatric speciation suggested by monophyly of crater lake cichlids. *Nature* 368: 629-632.

Seber, G.A.F. (1965) A note on the multiple-recapture census. *Biomerika* 52: 249.

Seber, G.A.F. (1973) *The Estimation of Animal Abundance and Related Parameters.* London: Griffin.

Seidel, B. (1987) Breeding of a *bombina variegata* population in a habitat with temporary pools. In J.J. Gelder (ed.), *Proceedings of the Fourth Ordinary General Meeting of the Societas Europaea Herpetologica.* Nijmegen.

Shaw, C.R.& Prasad, R. (1970) Starch gel electrophoresis of enzymes - a compilation of recipes. *Biochem.Genet.* 4: 297-320.

Shaw, D.D., Coates, D.J., Arnold, M.L.& Wilkinson, P. (1985) Temporal variation in the chromosomal structure of a hybrid zone and its relationship to karyotype repatterning. *Heredity* 55: 293-307.

Sheppard, P.M., Turner, J.R.G., Brown, K.S., Benson, W.W.& Singer, M.C. (1985) Genetics and the evolution of Muellerian mimicry in *Heliconius* butterflies. *Phil.Trans.Roy.Soc.Lond. B* 308: 433-613.

Simpson, G.G. (1961) *Principles of Animal Taxonomy*. New York: Columbia University Press.

Sites, J.W., Barton, N.H.& Reed, K.M. (in press) The Genetic structure of a mosaic hybrid zone between two chromosome race of the *Sceloporus grammicus* complex (Sauria, Phrynosomatidae) in central Mexico. *Evolution*.

Sites, J.W., Davis, S.K., Hutchison, D.W., Maurer, B.A.& Lara, G. (1993) Parapatric hybridization between chromosome races of the *Sceloporus grammicus* complex (Phrynosomatidae) : structure of the Tulancingo transect. *Copeia* 1993: (in press).

Slatkin, M. (1973) Gene flow and selection in a cline. *Genetics* 75: 733-756.

Slatkin, M. (1975) Gene flow and selection in a two-locus system. *Genetics* 81: 209-222.

Slatkin, M. (1985) Gene flow in natural populations. *Ann. Rev. Ecol. Syst.* 16: 393-430.

Southwood, T.R.E. (1978) *Ecological Methods.* Second Ed. London: Chapman and Hall.

Stebbins, G.L. (1950) *Variation and Evolution in Plants*. New York: Columbia University Press.

Szymura, J.M. (1976a) Hybridisation between discoglossid toads *Bombina bombina* and *B variegata* in southern Poland as revealed by the electrophoretic technique. *Z.Zool.Syst.Zool.Forsch.* 14: 227-236.

Szymura, J.M. (1976b) New data on the hybrid zone between *Bombina bombina* and *B variegata* (Anura: Discoglossidae). *Bull.Acad.Polon.Sci.Cl. II* 24: 355-363.

Szymura, J.M. (1983) Genetic differentiation between hybridising species *Bombina bombina* (L.) and *Bombina variegata* (L.) (Salkientia, Discoglossidae) in Poland. *Amphibia Reptilia* 4: 137-145.

Szymura, J.M. (1988) Regional differentiation and hybrid zones between fire-bellied toads *Bombina bombina* (L.) and *B. variegata* (L.) in Europe [in Polish]. Jagellonian University, Cracow.

Szymura, J.M. (1993) Analysis of hybrid zones with Bombina. In R.G. Harrison (ed.), *Hybrid zones and the Evolutionary Process*: 261-289. New York: Oxford University Press.

Szymura, J.M.& Barton, N.H. (1986) Genetic analysis of a hybrid zone between the fire-bellied toads *Bombina bombina* and B *variegata*, near Cracow in Southern Poland. *Evolution* 40: 1141-1159.

Szymura, J.M.& Barton, N.H. (1991) The genetic structure of the hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*: comparisons between transects and between loci. *Evolution* 45: 237-261.

Szymura, J.M.& Farana, I. (1978) Inheritance and linkage analysis of five enzyme loci in interspecific hybrids of toadlets, genus *Bombina*. *Biochem.Genet.* 16: 307-319.

Tauber, C.A.& Tauber, M.J. (1989) Sympatric speciation in insects: perception and perspective. In D.Otte & J. Endler (eds.), *Speciation and its consequences*: 307-344. Sunderland, Massachusetts: Sinauer Associates, Inc.

Templeton, A. (1989) The meaning of species and speciation. In D.Otte & J. Endler (eds.), *Speciation and its consequences*: 3-27. Sunderland, Massachusetts: Sinauer Associates, Inc.

Templeton, A.R. (1980) The theory of speciation via the founder principle. *Genetics* 94: 1011-1038.

Travis, J. (1984) Anuran size at metamorphosis: experimental test of a model based on intraspecific competition. *Ecology* 65: 1155-1160.

Turner, J.R.G. (1971) Müllerian mimicry in burnet moths and Heliconiid butterflies. In E.R Creed (ed.), *Ecological Genetics and Evolution*: 224-60. Oxford: Blackwell Scientific Publications.

Wahlund, S. (1928) Zusammensetzung von Population und Korrelationser scheinung vom Standpunkt der Vererbungslehre aus Betrachtet. *Hereditas* 11: 65-105.

Wecker, S.C. (1963) The role of early experience in habitat selection by the prairie deermouse *Peromyscus maniculatus bairdi*. *Ecol. Monogr* 33: 307-325.

White, M.J.D. (1978) *Modes of speciation*. W. H. Freeman (ed.), San Francisco.

Wilson, D.S.& Turelli, M. (1986) Stable underdominance and the evolutionary invasion of empty niches. *Amer.Nat.* 127: 835-850.

Wright, S. (1922) Coefficients of inbreeding and relationship. *Amer.Nat.* 56: 330-338.

Wright, S. (1931) Evolution in Mendelian populations. *Genetics* 16: 97-159.

Wright, S. (1932) The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proc.Sixth Int.Cong.Genet.* 1: 356-366.

Wright, S. (1940) The statistical consequences of Mendelian heredity in relation to speciation. In J.S. Huxley (ed.), *The New Systematics:* 161-183. Oxford: Clarendon press.

Wright, S. (1967) "Surfaces" of selective value. *Proc.Nat.Acad.Sci.(USA)* 58: 165-172.

Wright, S. (1978) *Evolution and the Genetics of Populations IV. Variability within and among Populationa.* Chicago: University of Chicago Press.

APPENDIX 2

The number of *bombina* (b) or *variegata* (v) alleles at each locus scored for each individual using horizontal starch gel electrophoresis. The first column is the site where the individual was found, the second column is the name of that individual and the following six columns are the number of b and v alleles at six allozymes. The allozymes are; adenylate kinase (AK), glucose phosphate isomerase (GPI), malate dehydrogenase (MDH), lactate dehydrogenase (LDH), isocitrate dehydrogenase (IDH) and creatine kinase (CK). A * represents missing data.

		AK	GPI	MDH	LDH	IDH	Ck
Site	IND	bv	bv	bv	bv	bv	bv
1	101	20	20	20	20	**	11
1	102	20	02	20	20	**	11
1	103	20	02	20	20	**	20
1	104	20	02	20	20	**	20
1	105	20	11	20	02	**	20
1	106	20	11	20	20	**	00
1	107	20	02	20	20	**	11
1	108	20	11	20	11	**	11
1	109	20	02	20	20	**	00
1	110	20	11	20	20	**	02
1	111	20	11	20	20	**	20
1	112	20	11	20	20	**	20
1	113	20	11	20	20	**	20
1	114	20	11	20	20	**	20
1	115	20	02	20	20	**	20
1	116	20	02	20	20	**	20
1	117	20	02	20	20	**	20
1	118	20	11	20	11	**	20
1	119	11	02	20	20	**	20
1	120	20	02	20	20	**	20
1	121	20	11	20	11	**	11
1	122	20	02	20	20	**	20
1	123	20	02	20	20	**	20
1	124	11	11	20	20	**	11
1	125	20	02	20	20	**	20
1	126	20	11	20	20	**	20
1	127	20	02	20	11	**	20
1	128	20	02	20	11	**	20
1	129	20	02	20	20	**	20
1	130	11	11	20	20	**	11
1	131	20	11	20	11	**	11
1	132	20	11	20	20	**	11
1	132	20	11	20	20	**	11
1	133	20	11	20	20	**	11
1	134	20	02	20	20	**	20
1	135	20	02	20	20	**	00
1	136	20	02	20	20	**	00
1	137	20	20	20	20	**	20
1	138	11	02	20	20	**	11
1	139	11	02	20	20	**	20
1	140	20	11	20	11	**	20
1	141	20	11	20	11	**	20
1	142	20	02	20	20	**	20
1	143	20	11	20	20	**	20
1	144	20	11	20	20	**	20
1	145	20	20	20	20	**	20
1	146	20	02	20	20	**	20
1	147	20	02	20	11	**	20
1	148	20	02	20	20	**	20
1	149	11	02	20	20	**	20
2	201	20	02	20	20	**	20
2	202	20	02	20	20	**	11
2	203	20	11	20	20	**	20
2	204	20	11	20	11	**	20
2	205	20	02	20	20	**	20
2	206	11	02	20	20	**	20
2	207	20	02	20	20	**	11
2	208	20	11	20	20	**	11
2	209	11	02	20	20	**	20
2	210	20	02	20	20	**	20
2	211	20	11	20	20	**	20
2	212	11	11	20	20	**	20
2	213	20	11	20	20	**	20
2	214	20	11	20	20	**	20
2	215	20	11	20	20	**	02
2	216	20	02	20	20	**	20
2	217	20	02	20	11	**	20
2	218	20	11	20	20	**	20
2	219	20	11	20	20	**	20
2	220	00	11	20	20	**	02
2	221	20	02	20	20	**	20
2	222	20	02	20	20	**	20
2	223	20	02	20	20	**	20
2	224	20	02	20	20	**	11
2	225	20	02	20	20	**	20
2	226	20	02	20	20	**	20
2	227	20	11	20	20	**	20
2	228	20	02	20	20	**	20
2	229	20	02	20	20	**	20
2	230	20	11	20	11	**	11
2	231	20	11	20	20	**	11
2	232	20	11	20	20	**	20
2	233	20	02	20	20	**	20
2	234	11	11	20	20	**	11
2	235	11	11	20	20	**	11
2	236	20	02	20	20	**	11
2	237	20	02	20	20	**	20
2	238	20	11	20	20	**	20
2	239	11	02	20	20	**	20
2	240	20	02	20	20	**	11
2	241	20	02	20	20	**	20
2	242	20	02	20	20	**	20
2	243	20	02	20	20	**	20
2	244	20	02	20	20	**	20
2	245	02	02	20	20	**	20
2	246	20	11	20	20	**	20
2	247	20	02	20	20	**	20

Site	IND	AK bv	GPI bv	MDH bv	LDH bv	IDH bv	Ck bv
2	248	20	11	20	20	**	20
2	249	20	02	20	20	**	11
3	301	20	02	20	20	**	20
3	302	20	20	20	20	**	20
3	303	20	02	20	20	**	20
3	304	20	02	20	20	**	20
3	305	20	11	20	11	**	20
3	306	20	02	20	20	**	20
3	307	20	02	20	20	**	20
3	308	20	02	20	20	**	20
3	309	20	11	20	11	**	20
3	310	20	02	20	20	**	20
3	311	20	11	20	20	**	11
3	312	11	02	20	11	**	11
3	313	20	11	20	20	**	20
3	314	20	02	20	20	**	11
3	315	20	02	20	20	**	20
3	316	20	02	20	20	**	20
3	317	11	02	20	11	**	20
3	318	20	02	20	20	**	20
3	319	20	02	20	20	**	20
4	401	20	02	20	20	**	20
4	402	20	11	20	20	**	20
4	403	20	11	20	20	**	20
4	404	20	02	20	20	**	11
4	405	20	02	20	20	**	20
4	406	11	02	20	20	**	20
4	407	20	02	20	20	**	11
4	408	11	02	20	20	**	20
4	409	20	02	20	11	**	20
4	410	20	02	20	20	**	20
4	411	20	02	20	20	**	20
4	412	20	02	20	20	**	11
4	413	11	02	20	20	**	11
4	414	20	11	20	20	**	11
4	415	20	02	20	11	**	20
4	416	20	02	20	20	**	20
4	417	20	02	20	11	**	20
4	418	20	02	20	20	**	20
4	419	20	02	20	11	**	20
4	420	20	02	20	20	**	20
4	421	20	02	20	20	**	20
4	422	20	02	20	20	**	20
5	501	02	02	20	20	**	20
5	502	20	02	20	20	**	20
5	503	20	02	20	20	**	20
5	504	20	02	20	20	**	20
5	505	20	02	20	20	**	20
5	506	20	02	20	20	**	20
5	507	20	02	20	11	**	20
5	508	20	11	20	11	**	20
5	509	11	02	20	20	**	11
5	510	20	11	20	20	**	20
5	511	20	02	20	20	**	20
5	512	20	02	20	20	**	20
5	513	20	02	20	20	**	11
5	514	20	02	20	11	**	11
5	515	20	02	20	20	**	20
5	516	11	02	20	20	**	20
5	517	20	11	20	11	**	20
5	518	20	02	20	20	**	11
5	519	11	02	20	20	**	20
5	520	20	02	11	11	**	11
5	521	20	02	11	11	**	20
5	522	20	02	20	20	**	20
5	523	11	02	11	20	**	20
5	524	11	02	20	20	**	20
5	525	11	02	20	20	**	20
5	526	20	02	20	20	**	11
5	527	11	02	20	02	**	20
5	528	20	02	20	20	**	20
5	529	11	11	20	20	**	11
6	601	11	02	20	20	**	11
6	602	20	02	20	11	**	20
6	603	20	11	20	20	**	20
6	604	20	11	20	20	**	11
6	605	20	11	20	20	**	20
6	606	20	02	20	20	**	11
6	606	20	02	20	20	**	11
6	608	20	02	20	20	**	20
6	609	20	02	20	20	**	20
6	610	02	02	20	20	**	11
6	611	20	11	20	20	**	20
6	612	20	02	20	11	**	20
6	613	20	11	20	11	**	20
6	614	20	02	11	20	**	11
6	615	11	02	11	20	**	11
6	616	11	02	20	11	**	20
6	617	11	02	20	20	**	20
6	618	20	02	20	20	**	20
6	619	11	11	20	20	**	20
6	620	20	02	20	20	**	20
6	621	20	11	20	11	**	20
6	622	02	02	20	11	**	20
6	623	20	02	20	20	**	11
6	624	20	02	20	20	**	11
6	625	20	02	20	20	**	20
6	626	20	02	20	11	**	11
6	627	20	11	20	20	**	20
6	628	20	20	20	20	**	20
6	629	20	02	20	20	**	20
6	630	11	02	20	11	**	11
6	631	20	02	20	20	**	20
6	632	20	02	20	20	**	11
6	633	20	02	20	11	**	20
6	634	20	02	20	20	**	11
6	635	20	20	20	20	**	20
7	701	02	02	02	11	**	11
7	702	20	02	11	20	**	02
7	703	02	02	11	02	**	11
7	704	11	02	11	11	**	20
7	705	20	02	20	20	**	20
7	706	11	02	02	02	**	02
7	707	20	02	02	02	**	11
7	708	11	02	11	11	**	11
7	709	20	02	11	02	**	11
8	801	20	02	02	02	**	02
8	802	02	02	02	02	**	02
8	803	02	02	02	02	**	02

Site	IND	AK bv	GPI bv	MDH bv	LDH bv	IDH bv	Ck bv
8	804	11	02	02	11	**	20
8	805	11	02	02	02	**	02
8	806	11	02	02	02	**	02
8	807	02	02	02	02	**	02
8	808	11	02	11	11	**	11
8	809	20	02	02	02	**	02
8	810	20	02	02	02	**	02
9	901	11	02	11	11	**	11
9	902	02	02	02	02	**	02
9	903	02	02	02	02	**	02
9	904	11	02	11	02	**	11
9	905	02	02	11	02	**	02
10	1001	02	02	11	20	**	20
10	1002	02	02	20	20	**	20
10	1003	11	02	20	11	**	20
10	1004	20	02	20	20	**	20
10	1005	11	02	11	11	**	02
10	1006	11	02	11	20	**	11
10	1007	20	02	20	20	**	11
11	1101	20	02	20	20	**	20
11	1102	20	11	20	11	**	20
11	1103	20	11	20	20	**	20
12	1201	02	02	11	02	**	02
12	1202	11	02	20	11	**	11
12	1203	11	11	02	11	**	11
12	1204	02	02	11	11	**	02
12	1205	02	02	11	02	**	20
13	1301	11	02	02	02	**	02
13	1302	20	02	11	20	**	11
13	1303	11	02	20	11	**	02
13	1304	20	11	11	11	**	11
13	1305	02	11	20	02	**	02
13	1306	02	02	11	02	**	02
14	1406	20	02	11	11	**	02
14	1407	20	11	20	11	**	20
14	1408	11	20	20	11	**	11
14	1409	11	02	11	02	**	02
14	1410	02	02	11	11	**	20
14	1411	11	02	11	11	**	20
14	1412	11	02	11	02	**	11
14	1413	02	02	02	11	**	02
14	1414	02	02	02	02	**	11
14	1415	02	02	02	02	**	20
14	1416	02	11	02	02	**	20
14	1417	02	02	11	20	**	00
15	1501	02	02	20	02	**	02
15	1502	11	02	02	02	**	02
15	1503	11	02	20	02	**	02
15	1504	02	02	02	02	**	02
15	1505	02	02	02	02	**	02
15	1506	02	02	02	02	**	02
15	1507	02	02	02	02	**	02
15	1508	02	02	02	02	**	02
15	1509	02	11	02	02	**	02
15	1510	02	02	02	02	**	11
15	1511	02	02	11	02	**	02
15	1512	02	02	02	02	**	02
15	1513	02	02	02	02	**	02
15	1514	02	02	02	02	**	02
15	1515	02	02	02	11	**	02
15	1516	02	02	02	02	**	02
15	1517	02	02	11	11	**	02
15	1518	02	02	02	02	**	02
15	1519	02	02	02	02	**	02
15	1520	11	02	02	02	**	02
15	1521	02	02	02	11	**	02
15	1522	02	02	02	02	**	02
15	1523	02	02	02	02	**	02
15	1524	02	02	11	02	**	02
15	1525	02	02	02	02	**	02
15	1526	11	02	02	02	**	02
15	1527	02	02	02	02	**	02
15	1528	02	02	02	02	**	11
15	1529	02	11	02	11	**	02
15	1530	02	02	02	02	**	02
15	1531	11	02	02	02	**	02
15	1532	02	02	02	02	**	02
15	1533	02	02	11	02	**	02
15	1534	02	02	02	02	**	02
15	1535	02	02	02	02	**	02
15	1536	02	02	02	02	**	02
15	1537	02	02	02	02	**	02
15	1538	02	02	02	02	**	02
15	1539	02	02	11	02	**	02
15	1540	02	02	02	02	**	02
15	1541	02	02	02	02	**	02
15	1542	02	02	02	02	**	02
15	1543	02	02	11	02	**	02
15	1544	02	02	02	02	**	02
15	1545	02	02	11	02	**	11
15	1546	02	02	11	02	**	02
15	1547	02	02	02	02	**	11
15	1548	02	02	11	02	**	02
15	1549	02	02	02	02	**	11
15	1550	02	02	02	02	**	02
15	1551	02	02	02	02	**	02
15	1552	11	02	02	11	**	02
15	1553	02	02	02	02	**	02
15	1554	02	02	02	02	**	02
15	1555	02	02	02	02	**	02
15	1556	02	02	20	02	**	02
15	1557	02	11	11	11	**	02
15	1558	02	02	02	02	**	02
15	1559	02	11	02	02	**	02
15	1560	02	02	02	02	**	02
15	1561	02	02	02	02	**	02
15	1562	02	11	02	02	**	02
15	1563	02	02	11	02	**	11
15	1564	02	11	02	02	**	02
16	1601	02	02	02	02	**	11
16	1602	11	02	11	02	**	11
16	1603	11	02	11	02	**	11
16	1604	02	02	02	02	**	11
16	1605	11	02	20	11	**	20
16	1606	02	02	02	02	**	11
16	1607	02	02	02	02	**	11
16	1608	02	02	02	11	**	02
16	1609	11	02	11	02	**	11
16	1610	02	02	02	02	**	02

Site	IND	AK bv	GPI bv	MDH bv	LDH bv	IDH bv	Ck bv
16	1611	02	02	11	02	**	02
16	1612	02	02	11	02	**	02
16	1613	11	02	02	02	**	02
16	1614	02	02	02	02	**	02
16	1615	02	02	20	11	**	11
16	1616	02	02	11	02	**	02
17	1701	02	02	02	02	**	02
17	1702	02	02	02	02	**	11
17	1703	02	02	02	02	**	02
17	1704	02	02	11	02	**	02
17	1705	02	02	02	11	**	02
17	1706	02	02	02	02	**	02
17	1707	02	02	02	02	**	02
17	1708	02	02	02	02	**	02
17	1709	02	02	02	02	**	02
17	1710	02	02	02	02	**	02
17	1711	02	02	02	02	**	02
17	1712	02	02	02	11	**	11
17	1713	02	02	02	02	**	02
17	1714	11	02	02	02	**	02
17	1715	02	02	02	02	**	02
17	1716	02	02	02	02	**	11
17	1717	11	02	02	02	**	02
17	1718	11	11	02	11	**	02
17	1719	02	02	02	02	**	02
17	1720	02	02	11	02	**	02
17	1721	02	02	02	11	**	02
17	1722	02	02	02	02	**	11
17	1723	02	02	02	02	**	02
17	1724	02	02	02	02	**	02
17	1725	02	02	02	11	**	02
17	1726	02	02	02	02	**	02
17	1727	02	02	02	02	**	11
17	1728	02	02	02	02	**	11
17	1729	02	02	02	02	**	11
17	1730	02	02	02	20	**	02
17	1731	02	02	11	02	**	02
17	1732	02	02	02	02	**	02
17	1733	02	02	02	02	**	02
17	1734	02	02	02	02	**	11
17	1735	02	02	02	11	**	11
18	1801	02	02	02	02	**	11
18	1802	11	02	02	11	**	11
18	1803	02	02	02	02	**	02
18	1804	11	02	11	11	**	02
18	1805	02	02	11	02	**	02
18	1806	02	02	02	02	**	02
18	1807	02	02	11	02	**	11
18	1808	02	02	02	02	**	02
18	1809	02	02	02	11	**	02
18	1810	02	02	02	02	**	02
18	1811	02	02	02	02	**	02
18	1812	11	02	02	02	**	02
18	1813	02	02	02	02	**	11
18	1814	02	02	11	02	**	02
18	1815	02	02	02	02	**	02
18	1816	02	02	02	02	**	02
18	1817	02	02	02	02	**	11
18	1818	02	02	02	02	**	11
18	1819	02	02	02	02	**	02
18	1820	02	02	02	02	**	02
18	1821	02	02	02	02	**	02
18	1822	02	02	11	02	**	02
18	1823	11	02	02	02	**	02
18	1824	02	02	02	11	**	02
18	1825	11	02	11	11	**	11
19	1901	02	02	02	02	**	02
19	1902	02	02	02	02	**	02
19	1903	02	02	02	02	**	11
19	1904	02	02	02	02	**	11
19	1905	02	02	11	02	**	02
19	1906	02	02	02	02	**	11
19	1907	02	02	02	02	**	02
19	1908	02	02	02	02	**	02
19	1909	02	02	02	02	**	02
19	1910	02	02	02	11	**	02
19	1911	02	02	02	02	**	02
19	1912	02	02	11	02	**	02
19	1913	02	02	11	02	**	02
19	1914	11	02	02	02	**	02
19	1915	02	02	20	02	**	02
19	1916	02	02	11	02	**	02
19	1917	02	02	02	02	**	11
19	1918	02	02	02	02	**	11
19	1919	02	02	11	02	**	02
19	1920	02	02	02	02	**	02
19	1921	11	02	02	02	**	02
19	1922	02	02	02	02	**	11
19	1923	02	02	02	02	**	02
19	1924	02	02	02	02	**	02
19	1925	02	02	02	02	**	11
19	1926	02	02	02	02	**	11
19	1927	02	02	11	02	**	02
19	1928	02	02	02	02	**	02
19	1929	11	02	02	02	**	02
19	1930	02	02	02	02	**	02
19	1931	02	02	02	11	**	11
19	1932	02	02	02	02	**	11
19	1933	02	02	02	02	**	02
19	1934	02	02	02	02	**	02
19	1935	02	02	02	02	**	02
19	1936	02	02	02	02	**	11
19	1937	11	02	02	02	**	02
19	1938	02	02	02	02	**	02
19	1939	02	02	11	02	**	11
19	1940	02	02	02	02	**	02
19	1941	02	02	02	02	**	02
19	1942	02	02	11	02	**	11
19	1943	11	02	02	02	**	02
19	1944	02	02	02	02	**	02
19	1945	02	02	02	02	**	02
19	1946	02	02	02	02	**	02
20	2001	02	02	02	02	**	02
20	2002	02	02	02	02	**	11
20	2003	02	02	11	02	**	11
20	2004	11	02	02	02	**	02
20	2005	02	02	02	02	**	02
20	2006	02	02	02	02	**	02
20	2007	02	02	02	02	**	11

Site	IND	AK bv	GPI bv	MDH bv	LDH bv	IDH bv	Ck bv				
20	2008	02	11	02	02	* *	02				
20	2009	02	02	02	02	* *	02				
20	2010	02	02	02	02	* *	02				
20	2011	02	02	02	02	* *	02				
20	2012	02	02	02	02	* *	02				
20	2013	02	02	02	02	* *	02				
20	2014	02	02	02	02	* *	02				
1001	1001000	0	2	0	2	1	1	0	2	2	0	**
1001	1001002	1	1	0	2	1	1	1	1	0	2	**
1001	1001003	0	2	0	2	0	2	1	1	1	1	**
1001	1001004	0	2	0	2	0	2	0	2	0	2	**
1001	1001005	1	1	0	2	1	1	1	1	0	2	**
1001	1001006	1	1	0	2	1	1	0	2	0	2	**
1001	1001007	0	2	0	2	1	1	0	2	0	2	**
1001	1001010	0	2	1	1	0	2	1	1	1	1	**
1001	1001011	1	1	0	2	1	1	0	2	1	1	**
1001	1001012	0	2	0	2	1	1	0	2	0	2	**
1001	1001013	0	2	0	2	1	1	0	2	0	2	**
1001	1001014	0	2	0	2	0	2	2	0	1	1	**
1001	1001016	1	1	0	2	0	2	1	1	*	*	**
1001	1001017	0	2	0	2	0	2	0	2	0	2	**
1001	1001018	0	2	1	1	1	1	2	0	0	2	**
1001	1001019	0	2	0	2	0	2	0	2	0	2	**
1001	1001020	1	1	0	2	2	0	2	0	2	0	**
1001	1001021	*	*	*	*	*	*	*	*	*	*	**
1001	1001022	*	*	*	*	*	*	*	*	*	*	**
1001	1001023	*	*	*	*	*	*	*	*	*	*	**
1001	1001024	*	*	*	*	*	*	*	*	*	*	**
1001	1001025	0	2	0	2	1	1	1	1	0	2	**
1001	1001026	0	2	0	2	0	2	0	2	0	2	**
1001	1001027	0	2	0	2	1	1	0	2	0	2	**
1001	1001028	1	1	0	2	1	1	0	2	0	2	**
1001	1001029	0	2	0	2	0	2	0	2	0	2	**
1001	1001030	0	2	0	2	0	2	0	2	2	0	**
1001	1001032	1	1	0	2	0	2	0	2	0	2	**
1001	1001033	2	0	1	1	0	2	1	1	2	0	**
1001	1001035	2	0	0	2	0	2	2	0	2	0	**
1001	1001044	0	2	0	2	1	1	0	2	0	2	**
1001	1001045	0	2	0	2	1	1	2	0	2	0	**
1002	1002001	0	2	1	1	0	2	0	2	1	1	**
1002	1002002	1	1	0	2	1	1	1	1	0	2	**
1002	1002003	1	1	0	2	0	2	0	2	0	2	**
1002	1002004	1	1	0	2	0	2	1	1	0	2	**
1002	1002005	1	1	0	2	1	1	0	2	0	2	**
1002	1002006	0	2	0	2	0	2	1	1	2	0	**
1002	1002007	0	2	0	2	0	2	0	2	0	2	**
1002	1002008	0	2	0	2	0	2	1	1	0	2	**
1002	1002009	0	2	0	2	1	1	0	2	0	2	**
1002	1002010	0	2	0	2	0	2	1	1	0	2	**
1002	1002013	1	1	0	2	2	0	1	1	1	1	**
1002	1002014	0	2	0	2	1	1	0	2	0	2	**
1002	1002015	1	1	0	2	0	2	1	1	0	2	**
1002	1002016	1	1	0	2	0	2	1	1	2	0	**
1002	1002017	1	1	1	1	0	2	0	2	2	0	**
1002	1002018	*	*	2	0	0	2	1	1	1	1	**
1002	1002019	2	0	0	2	2	0	1	1	2	0	**
1002	1002020	1	1	0	2	0	2	0	2	2	0	**
1003	1003001	0	2	0	2	2	0	0	2	1	1	**
1003	1003003	2	0	0	2	2	0	2	0	*	*	**
1003	1003004	0	2	0	2	1	1	0	2	1	1	**
1003	1003005	0	2	0	2	0	2	0	2	0	2	**
1003	1003006	1	1	0	2	0	2	1	1	1	1	**
1003	1003007	0	2	0	2	1	1	1	1	0	2	**
1003	1003008	2	0	0	2	0	2	0	2	0	2	**
1003	1003009	2	0	0	2	2	0	1	1	0	2	**
1003	1003010	1	1	0	2	1	1	0	2	1	1	**
1003	1003011	1	1	*	*	0	2	0	2	0	2	**
1003	1003012	0	2	*	*	0	2	1	1	1	1	**
1003	1003013	1	1	*	*	1	1	0	2	0	2	**
1003	1003014	0	2	*	*	0	2	1	1	0	2	**
1003	1003015	1	1	*	*	0	2	0	2	0	2	**
1003	1003016	1	1	0	2	0	2	0	2	0	2	**
1003	1003031	*	*	*	*	*	*	*	*	*	*	**
1003	1003032	*	*	*	*	*	*	*	*	*	*	**
1003	1003033	*	*	*	*	*	*	*	*	*	*	**
1003	1003034	*	*	*	*	*	*	*	*	*	*	**
1004	1004001	2	0	0	2	2	0	2	0	2	0	**
1004	1004002	0	2	1	1	0	2	0	2	0	2	**
1004	1004003	1	1	1	1	2	0	1	1	*	*	**
1005	1005001	2	0	0	2	2	0	1	1	2	0	**
1010	1010001	2	0	0	2	2	0	0	2	2	0	**
1013	1013001	2	0	1	1	2	0	2	0	2	0	**
1013	1013002	1	1	0	2	2	0	2	0	2	0	**
1013	1013003	2	0	0	2	2	0	2	0	2	0	**
1013	1013004	2	0	0	2	2	0	2	0	2	0	**
1013	1013005	2	0	0	2	2	0	2	0	2	0	**
1014	1014001	1	1	0	2	2	0	2	0	2	0	**
1014	1014002	2	0	0	2	1	1	2	0	2	0	**
1014	1014003	2	0	0	2	2	0	2	0	2	0	**
1014	1014004	2	0	0	2	2	0	2	0	2	0	**
1014	1014005	2	0	0	2	2	0	2	0	2	0	**
1014	1014006	2	0	0	2	2	0	2	0	2	0	**
1014	1014007	2	0	0	2	2	0	2	0	2	0	**
1014	1014008	2	0	0	2	2	0	2	0	2	0	**
1014	1014009	2	0	0	2	2	0	2	0	2	0	**
1014	1014010	2	0	0	2	2	0	2	0	2	0	**
1014	1014011	2	0	0	2	2	0	2	0	2	0	**
1015	1015001	0	2	0	2	1	1	0	2	0	2	**
1015	1015002	0	2	0	2	1	1	0	2	0	2	**
1015	1015003	1	1	0	2	1	1	1	1	1	1	**
1015	1015004	0	2	0	2	0	2	0	2	*	*	**
1015	1015005	0	2	0	2	0	2	0	2	*	*	**
1015	1015006	1	1	0	2	1	1	0	2	*	*	**
1015	1015007	0	2	0	2	0	2	1	1	*	*	**
1016	1016001	1	1	0	2	2	0	2	0	2	0	**
1016	1016002	2	0	0	2	2	0	2	0	2	0	**
1016	1016003	2	0	0	2	2	0	2	0	2	0	**
1018	1018001	0	2	0	2	2	0	0	2	2	0	**
1019	1019001	2	0	1	1	2	0	2	0	2	0	**
1019	1019002	2	0	0	2	2	0	2	0	2	0	**
1019	1019003	2	0	0	2	2	0	2	0	2	0	**
1019	1019004	*	*	*	*	*	*	*	*	*	*	**
1019	1019005	1	1	*	*	0	2	0	2	0	2	02
1019	1019006	1	1	*	*	2	0	2	0	2	0	20
1025	1025001	0	2	*	*	0	2	0	2	0	2	**
1028	1028008	0	2	*	*	0	2	0	2	0	2	02
1028	1028009	0	2	*	*	0	2	0	2	0	2	02
1028	1028010	0	2	*	*	0	2	0	2	0	2	02

Site	IND	AK b	AK v	GPI b	GPI v	MDH b	MDH v	LDH b	LDH v	IDH b	IDH v	Ck bv
1028	1028011	0	2	*	*	0	2	0	2	0	2	02
1028	1028012	0	2	*	*	0	2	0	2	0	2	02
1028	1028013	1	1	*	*	0	2	0	2	0	2	02
1029	1029001	0	2	*	*	0	2	0	2	0	2	02
1029	1029002	0	2	*	*	0	2	0	2	0	2	02
1029	1029003	0	2	*	*	0	2	0	2	0	2	02
1029	1029004	0	2	*	*	0	2	0	2	0	2	02
1029	1029005	0	2	*	*	0	2	0	2	0	2	02
1029	1029006	0	2	*	*	0	2	0	2	0	2	02
1029	1029007	0	2	*	*	0	2	0	2	0	2	02
1029	1029008	0	2	*	*	0	2	0	2	0	2	02
1029	1029009	0	2	*	*	0	2	0	2	0	2	02
1029	1029010	0	2	*	*	0	2	0	2	0	2	02
1029	1029011	0	2	*	*	0	2	0	2	0	2	02
1029	1029012	0	2	*	*	0	2	0	2	0	2	02
1029	1029013	0	2	*	*	0	2	0	2	0	2	02
1029	1029014	0	2	*	*	0	2	0	2	0	2	02
1029	1029015	0	2	0	2	0	2	0	2	0	2	**
1029	1029016	0	2	0	2	0	2	0	2	0	2	**
1029	1029017	0	2	0	2	1	1	0	2	0	2	**
1029	1029020	0	2	0	2	0	2	0	2	0	2	**
1029	1029021	0	2	0	2	0	2	0	2	0	2	**
1029	1029022	0	2	0	2	0	2	0	2	0	2	**
1029	1029023	1	1	0	2	0	2	0	2	0	2	**
1029	1029024	0	2	0	2	0	2	0	2	0	2	**
1029	1029025	0	2	0	2	0	2	0	2	0	2	**
1029	1029027	0	2	0	2	0	2	0	2	0	2	**
1029	1029028	0	2	0	2	0	2	0	2	0	2	**
1029	1029029	0	2	0	2	0	2	0	2	0	2	**
1029	1029030	0	2	0	2	0	2	0	2	0	2	**
1029	1029033	0	2	0	2	0	2	0	2	0	2	**
1029	1029035	0	2	0	2	0	2	0	2	*	*	**
1029	1029036	0	2	0	2	0	2	0	2	0	2	**
1029	1029037	0	2	0	2	0	2	0	2	0	2	**
1029	1029038	0	2	0	2	1	1	0	2	0	2	**
1029	1029039	0	2	0	2	0	2	0	2	0	2	**
1029	1029042	2	0	0	2	0	2	0	2	0	2	**
1029	1029044	0	2	0	2	0	2	0	2	0	2	**
1029	1029045	0	2	0	2	0	2	0	2	0	2	**
1029	1029046	0	2	0	2	0	2	0	2	0	2	**
1032	1032001	2	0	*	*	2	0	2	0	2	0	20
1033	1033001	2	0	*	*	2	0	2	0	2	0	20
1033	1033002	2	0	*	*	2	0	2	0	2	0	20
1035	1035001	1	1	0	2	2	0	2	0	*	*	**
1035	1035002	2	0	0	2	1	1	2	0	*	*	**
1035	1035003	1	1	1	1	2	0	2	0	*	*	**
1035	1035004	2	0	2	0	2	0	2	0	*	*	**
1035	1035005	1	1	0	2	2	0	1	1	2	0	**
1035	1035006	2	0	1	1	2	0	1	1	2	0	**
1035	1035007	2	0	1	1	2	0	2	0	2	0	**
1035	1035008	2	0	0	2	2	0	2	0	2	0	**
1035	1035009	2	0	0	2	2	0	2	0	*	*	**
1035	1035010	2	0	0	2	2	0	2	0	*	*	**
1035	1035011	2	0	0	2	0	2	2	0	*	*	**
1035	1035012	1	1	0	2	2	0	2	0	2	0	**
1035	1035013	2	0	0	2	2	0	2	0	2	0	**
1035	1035014	1	1	0	2	2	0	2	0	2	0	**
1035	1035015	2	0	0	2	2	0	2	0	2	0	**
1035	1035016	2	0	0	2	2	0	1	1	2	0	**
1035	1035017	2	0	2	0	2	0	2	0	2	0	**
1035	1035018	2	0	0	2	2	0	2	0	1	1	**
1035	1035019	1	1	1	1	2	0	2	0	2	0	**
1035	1035020	2	0	0	2	2	0	2	0	*	*	**
1035	1035021	1	1	0	2	2	0	2	0	*	*	**
1035	1035022	2	0	0	2	2	0	2	0	*	*	**
1035	1035023	1	1	0	2	2	0	1	1	*	*	**
1035	1035024	2	0	1	1	2	0	2	0	2	0	**
1035	1035025	2	0	0	2	2	0	2	0	2	0	**
1035	1035026	1	1	0	2	2	0	2	0	2	0	**
1035	1035027	1	1	1	1	2	0	2	0	2	0	**
1035	1035028	2	0	1	1	2	0	2	0	2	0	**
1035	1035029	2	0	0	2	2	0	2	0	2	0	**
1035	1035030	2	0	0	2	2	0	2	0	2	0	**
1035	1035031	1	1	0	2	2	0	2	0	2	0	**
1035	1035032	2	0	0	2	1	1	2	0	2	0	**
1035	1035033	2	0	0	2	2	0	2	0	2	0	**
1035	1035034	2	0	0	2	2	0	2	0	2	0	**
1036	1036001	2	0	0	2	2	0	1	1	2	0	**
1036	1036002	2	0	1	1	2	0	2	0	2	0	**
1037	1037001	2	0	*	*	2	0	2	0	2	0	20
1037	1037002	2	0	*	*	2	0	2	0	2	0	11
1038	1038001	1	1	*	*	2	0	2	0	2	0	20
1038	1038002	1	1	*	*	0	2	0	2	0	2	02
1038	1038003	1	1	*	*	2	0	2	0	2	0	20
1039	1039001	2	0	0	2	2	0	2	0	2	0	**
1039	1039002	2	0	1	1	2	0	1	1	2	0	**
1039	1039003	2	0	1	1	2	0	2	0	2	0	**
1039	1039004	2	0	1	1	2	0	1	1	2	0	**
1039	1039005	2	0	0	2	2	0	2	0	2	0	**
1039	1039006	0	2	0	2	2	0	2	0	1	1	**
1039	1039007	2	0	1	1	2	0	1	1	*	*	**
1039	1039008	2	0	0	2	2	0	2	0	*	*	**
1039	1039009	2	0	0	2	2	0	2	0	*	*	**
1039	1039013	2	0	0	2	2	0	2	0	2	0	**
1039	1039014	2	0	0	2	2	0	2	0	2	0	**
1039	1039015	1	1	0	2	2	0	2	0	2	0	**
1039	1039016	2	0	0	2	2	0	0	2	*	*	**
1039	1039017	2	0	0	2	2	0	1	1	*	*	**
1039	1039018	2	0	0	2	2	0	2	0	*	*	**
1039	1039020	2	0	0	2	2	0	2	0	2	0	**
1039	1039021	2	0	0	2	2	0	2	0	2	0	**
1039	1039031	2	0	0	2	2	0	1	1	2	0	**
1039	1039032	2	0	0	2	2	0	2	0	2	0	**
1039	1039033	2	0	0	2	1	1	2	0	2	0	**
1039	1039034	2	0	0	2	2	0	2	0	*	*	**
1039	1039035	1	1	0	2	2	0	2	0	*	*	**
1039	1039036	2	0	0	2	2	0	1	1	2	0	**
1039	1039037	2	0	0	2	2	0	1	1	*	*	**
1039	1039038	2	0	0	2	2	0	2	0	*	*	**
1039	1039039	2	0	0	2	2	0	2	0	*	*	**
1039	1039040	2	0	0	2	2	0	2	0	*	*	**
1039	1039041	2	0	0	2	2	0	1	1	2	0	**
1039	1039042	2	0	0	2	2	0	2	0	2	0	**
1039	1039043	2	0	0	2	2	0	2	0	2	0	**
1039	1039050	2	0	0	2	2	0	2	0	2	0	**
1039	1039051	2	0	0	2	2	0	2	0	2	0	**
1039	1039052	2	0	0	2	2	0	2	0	2	0	**
1039	1039053	2	0	0	2	2	0	2	0	1	1	**
1039	1039055	2	0	0	2	2	0	2	0	2	0	**

Site	IND	AK b	AK v	GPI b	GPI v	MDH b	MDH v	LDH b	LDH v	IDH b	IDH v	Ck bv
1039	1039056	2	0	0	2	2	0	2	0	1	1	**
1039	1039057	1	1	1	1	2	0	1	1	2	0	**
1039	1039059	2	0	0	2	2	0	2	0	2	0	**
1039	1039060	2	0	0	2	2	0	1	1	2	0	**
1039	1039061	2	0	0	2	2	0	2	0	2	0	**
1039	1039063	2	0	1	1	2	0	2	0	2	0	**
1039	1039064	1	1	0	2	2	0	1	1	2	0	**
1039	1039067	2	0	0	2	2	0	2	0	2	0	**
1039	1039068	2	0	0	2	2	0	2	0	2	0	**
1039	1039070	1	1	0	2	2	0	2	0	2	0	**
1039	1039071	2	0	0	2	2	0	2	0	2	0	**
1039	1039072	2	0	1	1	2	0	2	0	2	0	**
1039	1039074	2	0	0	2	2	0	2	0	2	0	**
1039	1039075	2	0	0	2	2	0	0	2	2	0	**
1039	1039078	2	0	0	2	2	0	2	0	2	0	**
1039	1039079	2	0	0	2	2	0	2	0	1	1	**
1039	1039080	2	0	2	0	2	0	2	0	2	0	**
1039	1039081	2	0	0	2	2	0	2	0	2	0	**
1039	1039082	2	0	1	1	2	0	2	0	2	0	**
1039	1039083	2	0	0	2	2	0	2	0	2	0	**
1040	1040001	1	1	0	2	2	0	2	0	2	0	**
1040	1040002	2	0	0	2	2	0	2	0	2	0	**
1040	1040003	2	0	0	2	2	0	1	1	2	0	**
1040	1040007	2	0	0	2	2	0	2	0	2	0	**
1040	1040008	1	1	0	2	2	0	1	1	2	0	**
1040	1040009	2	0	0	2	2	0	2	0	2	0	**
1040	1040010	2	0	0	2	2	0	1	1	2	0	**
1040	1040011	2	0	0	2	2	0	2	0	2	0	**
1040	1040012	2	0	0	2	2	0	2	0	2	0	**
1040	1040013	1	1	0	2	2	0	1	1	*	*	**
1040	1040014	2	0	0	2	2	0	2	0	*	*	**
1040	1040015	2	0	0	2	2	0	2	0	*	*	**
1040	1040016	2	0	0	2	2	0	2	0	*	*	**
1040	1040021	2	0	1	1	2	0	1	1	*	*	**
1040	1040022	1	1	0	2	2	0	2	0	0	2	**
1040	1040023	2	0	0	2	2	0	2	0	2	0	**
1040	1040025	2	0	0	2	2	0	2	0	2	0	**
1040	1040027	1	1	0	2	2	0	2	0	2	0	**
1040	1040028	2	0	1	1	2	0	2	0	2	0	**
1040	1040029	2	0	0	2	2	0	2	0	2	0	**
1040	1040034	2	0	1	1	2	0	2	0	2	0	**
1040	1040035	2	0	0	2	1	1	2	0	2	0	**
1040	1040036	2	0	1	1	2	0	0	2	2	0	**
1040	1040037	2	0	1	1	2	0	2	0	2	0	**
1040	1040038	2	0	0	2	2	0	1	1	2	0	**
1040	1040039	2	0	0	2	2	0	2	0	2	0	**
1040	1040040	2	0	1	1	2	0	2	0	2	0	**
1040	1040041	2	0	0	2	2	0	2	0	2	0	**
1040	1040047	2	0	1	1	2	0	2	0	2	0	**
1040	1040050	2	0	1	1	2	0	1	1	2	0	**
1040	1040051	2	0	1	1	2	0	2	0	2	0	**
1040	1040056	2	0	0	2	2	0	2	0	2	0	**
1040	1040057	2	0	1	1	2	0	2	0	2	0	**
1040	1040058	2	0	0	2	2	0	2	0	2	0	**
1040	1040059	2	0	0	2	2	0	2	0	2	0	**
1040	1040061	2	0	0	2	2	0	1	1	0	2	**
1041	1041001	1	1	*	*	0	2	0	2	1	1	20
1042	1042001	2	0	0	2	2	0	1	1	*	*	**
1042	1042002	2	0	0	2	2	0	2	0	*	*	**
1042	1042003	2	0	1	1	2	0	2	0	2	0	**
1042	1042004	2	0	0	2	2	0	2	0	*	*	**
1042	1042005	2	0	0	2	1	1	2	0	*	*	**
1042	1042006	2	0	1	1	2	0	2	0	2	0	**
1042	1042007	2	0	0	2	2	0	2	0	2	0	**
1042	1042008	1	1	0	2	1	1	2	0	2	0	**
1042	1042009	2	0	1	1	2	0	2	0	2	0	**
1042	1042010	2	0	0	2	2	0	2	0	2	0	**
1043	1043001	2	0	0	2	1	1	2	0	*	*	**
1043	1043002	2	0	0	2	2	0	2	0	*	*	**
1043	1043003	2	0	1	1	2	0	2	0	*	*	**
1043	1043004	1	1	0	2	2	0	1	1	*	*	**
1043	1043005	1	1	0	2	2	0	1	1	2	0	**
1043	1043006	2	0	2	0	2	0	2	0	2	0	**
1043	1043007	1	1	1	1	2	0	2	0	2	0	**
1043	1043008	2	0	0	2	2	0	1	1	2	0	**
1043	1043009	2	0	0	2	2	0	2	0	2	0	**
1043	1043010	2	0	0	2	2	0	2	0	2	0	**
1043	1043011	2	0	0	2	2	0	1	1	2	0	**
1043	1043013	2	0	1	1	2	0	2	0	1	1	**
1043	1043014	2	0	0	2	2	0	1	1	2	0	**
1043	1043015	2	0	0	2	2	0	1	1	2	0	**
1043	1043016	2	0	0	2	2	0	1	1	2	0	**
1043	1043017	2	0	1	1	1	1	2	0	2	0	**
1043	1043018	2	0	2	0	1	1	1	1	2	0	**
1043	1043019	1	1	0	2	2	0	0	2	*	*	**
1043	1043020	2	0	0	2	2	0	2	0	2	0	**
1043	1043021	1	1	0	2	2	0	1	1	2	0	**
1043	1043022	2	0	1	1	2	0	2	0	2	0	**
1043	1043023	2	0	0	2	2	0	1	1	1	1	**
1043	1043024	2	0	0	2	2	0	2	0	2	0	**
1043	1043025	2	0	*	*	2	0	2	0	2	0	**
1044	1044001	1	1	1	1	2	0	2	0	2	0	**
1044	1044002	2	0	1	1	1	1	0	2	2	0	**
1044	1044003	2	0	1	1	2	0	2	0	*	*	**
1044	1044004	2	0	0	2	1	1	1	1	2	0	**
1044	1044006	0	2	1	1	2	0	2	0	2	0	**
1044	1044007	1	1	2	0	1	1	0	2	2	0	**
1044	1044008	2	0	0	2	2	0	1	1	1	1	**
1044	1044009	2	0	0	2	2	0	2	0	2	0	**
1044	1044010	2	0	0	2	2	0	2	0	2	0	**
1044	1044011	1	1	0	2	2	0	1	1	1	1	**
1044	1044012	0	2	1	1	2	0	2	0	2	0	**
1044	1044017	2	0	0	2	1	1	1	1	2	0	**
1044	1044018	2	0	0	2	2	0	2	0	2	0	**
1044	1044019	2	0	0	2	2	0	0	2	0	2	**
1044	1044020	1	1	2	0	1	1	1	1	2	0	**
1044	1044021	2	0	0	2	2	0	1	1	*	*	**
1044	1044022	1	1	0	2	2	0	2	0	*	*	**
1044	1044023	2	0	0	2	2	0	2	0	*	*	**
1044	1044024	1	1	*	*	2	0	1	1	0	2	**
1045	1045001	2	0	0	2	2	0	1	1	2	0	**
1045	1045002	2	0	0	2	1	1	2	0	2	0	**
1045	1045003	2	0	0	2	2	0	2	0	2	0	**
1045	1045004	1	1	0	2	2	0	2	0	2	0	**
1045	1045005	2	0	0	2	2	0	1	1	2	0	**
1045	1045006	2	0	0	2	2	0	2	0	2	0	**
1045	1045007	1	1	0	2	2	0	2	0	2	0	**
1045	1045008	2	0	2	0	2	0	2	0	2	0	**
1045	1045009	2	0	*	*	2	0	2	0	2	0	**

Site	IND	AK b	AK v	GPI b	GPI v	MDH b	MDH v	LDH b	LDH v	IDH b	IDH v	Ck bv
1045	1045010	2	0	1	1	2	0	0	2	2	0	**
1046	1046003	0	2	*	*	1	1	1	1	2	0	**
1046	1046004	0	2	0	2	1	1	1	1	2	0	**
1046	1046005	1	1	0	2	2	0	2	0	0	2	**
1047	1047003	0	2	*	*	0	2	1	1	0	2	**
1047	1047004	2	0	*	*	2	0	2	0	2	0	**
1049	1049001	1	1	*	*	0	2	1	1	0	2	02
1049	1049002	2	0	*	*	2	0	2	0	1	1	11
1049	1049003	1	1	*	*	0	2	1	1	1	1	11
1049	1049004	1	1	*	*	0	2	0	2	1	1	11
1049	1049005	0	2	*	*	0	2	0	2	0	2	11
1050	1050001	2	0	*	*	2	0	2	0	2	0	20
1050	1050002	2	0	*	*	2	0	2	0	2	0	20
1050	1050003	2	0	*	*	2	0	2	0	2	0	20
1050	1050004	2	0	*	*	2	0	2	0	2	0	20
1050	1050005	2	0	*	*	2	0	2	0	2	0	20
1050	1050006	2	0	*	*	2	0	2	0	2	0	20
1050	1050007	2	0	*	*	2	0	2	0	2	0	20
1050	1050008	1	1	*	*	2	0	2	0	2	0	20
1050	1050009	2	0	*	*	2	0	2	0	2	0	20
1050	1050010	2	0	*	*	2	0	2	0	2	0	20
1050	1050011	2	0	*	*	2	0	2	0	2	0	20
1050	1050012	2	0	*	*	2	0	2	0	2	0	20
1050	1050013	2	0	*	*	2	0	2	0	2	0	20
1050	1050014	2	0	*	*	2	0	2	0	2	0	20
1050	1050015	2	0	*	*	2	0	2	0	2	0	20
1050	1050016	2	0	*	*	2	0	1	1	2	0	11
1050	1050017	1	1	*	*	2	0	2	0	1	1	20
1050	1050018	1	1	*	*	2	0	2	0	1	1	20
1051	1051001	1	1	*	*	2	0	2	0	2	0	02
1051	1051002	2	0	*	*	2	0	2	0	2	0	20
1052	1052001	2	0	*	*	2	0	2	0	2	0	20
1052	1052002	2	0	*	*	2	0	2	0	2	0	11
1052	1052003	2	0	*	*	2	0	2	0	2	0	20
1052	1052004	2	0	*	*	2	0	2	0	2	0	11
1052	1052005	2	0	*	*	2	0	1	1	2	0	20
1052	1052006	2	0	*	*	2	0	2	0	2	0	20
1052	1052007	2	0	*	*	2	0	2	0	2	0	20
1052	1052008	2	0	*	*	2	0	0	2	2	0	20
1052	1052009	2	0	*	*	2	0	2	0	2	0	20
1052	1052010	2	0	*	*	2	0	1	1	2	0	20
1052	1052011	0	2	*	*	2	0	2	0	2	0	20
1052	1052012	2	0	*	*	2	0	2	0	2	0	20
1052	1052016	2	0	0	2	2	0	2	0	2	0	**
1052	1052017	2	0	1	1	2	0	2	0	2	0	**
1052	1052018	2	0	0	2	2	0	2	0	2	0	**
1052	1052028	2	0	0	2	2	0	2	0	2	0	**
1052	1052029	2	0	0	2	2	0	2	0	2	0	**
1052	1052030	2	0	0	2	2	0	1	1	2	0	**
1052	1052031	1	1	1	1	2	0	2	0	2	0	**
1052	1052032	2	0	0	2	2	0	2	0	2	0	**
1052	1052033	2	0	0	2	2	0	2	0	2	0	**
1052	1052034	2	0	0	2	2	0	1	1	2	0	**
1052	1052035	2	0	0	2	2	0	0	2	2	0	**
1052	1052036	*	*	0	2	2	0	2	0	2	0	**
1052	1052040	2	0	0	2	2	0	0	2	2	0	**
1052	1052041	2	0	0	2	2	0	2	0	2	0	**
1052	1052042	1	1	0	2	2	0	2	0	2	0	**
1052	1052043	1	1	*	*	2	0	2	0	2	0	**
1052	1052044	2	0	*	*	2	0	2	0	2	0	**
1052	1052045	2	0	*	*	2	0	2	0	2	0	**
1052	1052049	2	0	2	0	2	0	2	0	*	*	**
1052	1052050	2	0	0	2	2	0	2	0	*	*	**
1052	1052051	2	0	1	1	2	0	2	0	*	*	**
1052	1052052	1	1	0	2	2	0	2	0	*	*	**
1052	1052053	2	0	0	2	2	0	1	1	2	0	**
1052	1052054	2	0	1	1	2	0	2	0	2	0	**
1052	1052055	2	0	0	2	2	0	2	0	2	0	**
1052	1052056	1	1	0	2	2	0	2	0	2	0	**
1053	1053006	2	0	0	2	2	0	1	1	2	0	**
1053	1053007	2	0	0	2	2	0	2	0	2	0	**
1053	1053008	2	0	0	2	2	0	2	0	2	0	**
1053	1053009	1	1	0	2	2	0	2	0	1	1	**
1053	1053010	2	0	2	0	2	0	2	0	2	0	**
1053	1053012	2	0	1	1	2	0	2	0	2	0	**
1053	1053014	2	0	1	1	2	0	2	0	2	0	**
1053	1053015	2	0	0	2	2	0	2	0	2	0	**
1053	1053016	2	0	0	2	2	0	1	1	2	0	**
1053	1053017	2	0	0	2	2	0	2	0	2	0	**
1053	1053018	2	0	0	2	2	0	2	0	2	0	**
1053	1053019	2	0	0	2	2	0	2	0	2	0	**
1053	1053020	1	1	0	2	2	0	2	0	2	0	**
1053	1053021	2	0	0	2	2	0	2	0	2	0	**
1053	1053022	2	0	0	2	2	0	2	0	2	0	**
1053	1053024	2	0	0	2	2	0	2	0	2	0	**
1053	1053027	2	0	0	2	2	0	2	0	2	0	**
1054	1054001	2	0	0	2	2	0	2	0	*	*	**
1054	1054002	1	1	0	2	2	0	1	1	*	*	**
1054	1054003	0	2	1	1	1	1	0	2	*	*	**
1054	1054004	2	0	1	1	1	1	0	2	*	*	**
1054	1054005	2	0	0	2	2	0	0	2	2	0	**
1054	1054006	0	2	0	2	2	0	1	1	1	1	**
1054	1054007	0	2	0	2	1	1	2	0	1	1	**
1054	1054008	2	0	0	2	1	1	0	2	0	2	**
1054	1054009	1	1	*	*	2	0	2	0	1	1	**
1054	1054010	0	2	*	*	0	2	0	2	0	2	**
1054	1054011	1	1	*	*	1	1	1	1	0	2	**
1054	1054012	1	1	*	*	1	1	1	1	0	2	**
1054	1054013	0	2	*	*	0	2	0	2	0	2	**
1054	1054014	1	1	*	*	0	2	1	1	0	2	**
1054	1054015	0	2	*	*	1	1	0	2	*	*	**
1054	1054016	1	1	*	*	0	2	0	2	*	*	**
1054	1054017	1	1	*	*	1	1	1	1	1	1	**
1054	1054018	0	2	*	*	2	0	1	1	1	1	**
1054	1054019	0	2	*	*	0	2	0	2	1	1	**
1054	1054020	0	2	*	*	0	2	0	2	0	2	**
1054	1054023	2	0	1	1	2	0	2	0	2	0	**
1054	1054026	*	*	*	*	*	*	*	*	*	*	**
1054	1054028	1	1	0	2	0	2	1	1	2	0	**
1054	1054029	1	1	0	2	1	1	0	2	2	0	**
1054	1054030	0	2	0	2	1	1	2	0	2	0	**
1054	1054031	1	1	0	2	2	0	1	1	0	2	**
1054	1054032	1	1	0	2	0	2	0	2	0	2	**
1054	1054033	0	2	0	2	0	2	1	1	1	1	**
1054	1054034	1	1	*	*	1	1	0	2	0	2	**
1055	1055002	2	0	1	1	2	0	2	0	2	0	**
1055	1055003	2	0	0	2	2	0	2	0	0	2	**
1055	1055004	0	2	0	2	0	2	1	1	0	2	**
1055	1055005	0	2	0	2	0	2	0	2	0	2	**

Site	IND	AK b	AK v	GPI b	GPI v	MDH b	MDH v	LDH b	LDH v	IDH b	IDH v	Ck bv
1055	1055006	2	0	0	2	2	0	1	1	1	1	**
1055	1055007	2	0	0	2	2	0	2	0	1	1	**
1055	1055008	2	0	0	2	2	0	2	0	2	0	**
1055	1055009	2	0	0	2	2	0	2	0	2	0	**
1055	1055010	1	1	0	2	1	1	1	1	2	0	**
1055	1055011	2	0	0	2	2	0	2	0	2	0	**
1055	1055012	2	0	0	2	2	0	2	0	2	0	**
1055	1055013	1	1	1	1	2	0	2	0	2	0	**
1055	1055014	2	0	0	2	2	0	1	1	2	0	**
1055	1055015	2	0	0	2	2	0	2	0	2	0	**
1055	1055016	2	0	0	2	2	0	2	0	2	0	**
1055	1055017	2	0	0	2	2	0	2	0	*	*	**
1055	1055018	2	0	0	2	1	1	2	0	*	*	**
1055	1055019	2	0	0	2	2	0	2	0	*	*	**
1055	1055020	2	0	0	2	2	0	2	0	2	0	**
1056	1056006	2	0	0	2	2	0	2	0	2	0	**
1056	1056007	1	1	0	2	0	2	0	2	1	1	**
1056	1056008	2	0	0	2	2	0	2	0	2	0	**
1056	1056009	2	0	*	*	2	0	1	1	2	0	**
1056	1056010	2	0	*	*	2	0	2	0	2	0	**
1056	1056011	2	0	*	*	2	0	2	0	2	0	**
1056	1056012	2	0	*	*	2	0	2	0	2	0	**
1056	1056013	1	1	0	2	2	0	2	0	*	*	**
1056	1056014	0	2	0	2	1	1	1	1	*	*	**
1056	1056015	1	1	2	0	2	0	1	1	*	*	**
1056	1056016	1	1	0	2	2	0	2	0	2	0	**
1056	1056017	2	0	0	2	2	0	2	0	2	0	**
1056	1056018	1	1	0	2	1	1	1	1	2	0	**
1056	1056019	2	0	0	2	2	0	2	0	2	0	**
1056	1056020	2	0	0	2	2	0	2	0	2	0	**
1056	1056021	1	1	0	2	1	1	2	0	0	2	**
1056	1056022	1	1	0	2	2	0	0	2	*	*	**
1056	1056023	0	2	0	2	0	2	0	2	*	*	**
1056	1056024	2	0	0	2	2	0	2	0	2	0	**
1057	1057001	2	0	2	0	2	0	2	0	*	*	**
1057	1057002	2	0	0	2	2	0	2	0	*	*	**
1057	1057003	2	0	2	0	2	0	2	0	*	*	**
1057	1057004	2	0	0	2	2	0	2	0	*	*	**
1058	1058001	0	2	*	*	0	2	0	2	0	2	02
1059	1059001	1	1	*	*	1	1	0	2	1	1	02
1059	1059002	0	2	*	*	0	2	0	2	1	1	02
1059	1059003	1	1	*	*	1	1	0	2	1	1	02
1060	1060001	1	1	0	2	1	1	0	2	0	2	**
1061	1061001	2	0	1	1	1	1	1	1	1	1	**
1061	1061002	2	0	0	2	2	0	2	0	2	0	**
1063	1063001	1	1	0	2	2	0	2	0	2	0	**
1063	1063002	2	0	0	2	2	0	2	0	2	0	**
1063	1063003	2	0	0	2	2	0	2	0	2	0	**
1063	1063004	0	2	0	2	2	0	1	1	2	0	**
1063	1063005	2	0	2	0	2	0	2	0	2	0	**
1063	1063006	1	1	0	2	2	0	1	1	0	2	**
1063	1063007	1	1	2	0	2	0	2	0	0	2	**
1063	1063008	0	2	0	2	2	0	1	1	2	0	**
1063	1063009	2	0	0	2	1	1	0	2	1	1	**
1063	1063010	1	1	1	1	1	1	1	1	0	2	**
1063	1063011	2	0	0	2	2	0	2	0	2	0	**
1063	1063012	0	2	0	2	1	1	2	0	1	1	**
1063	1063013	2	0	0	2	1	1	0	2	0	2	**
1063	1063014	2	0	1	1	2	0	1	1	2	0	**
1063	1063015	0	2	0	2	1	1	1	1	1	1	**
1063	1063016	1	1	0	2	2	0	2	0	0	2	**
1063	1063017	2	0	0	2	1	1	1	1	2	0	**
1063	1063018	2	0	0	2	2	0	2	0	2	0	**
1063	1063019	1	1	0	2	1	1	2	0	1	1	**
1063	1063020	1	1	0	2	2	0	1	1	1	1	**
1063	1063021	2	0	0	2	2	0	1	1	2	0	**
1063	1063022	1	1	0	2	1	1	1	1	1	1	**
1063	1063023	2	0	0	2	2	0	2	0	2	0	**
1063	1063024	1	1	1	1	2	0	2	0	2	0	**
1063	1063025	2	0	0	2	2	0	2	0	2	0	**
1063	1063026	0	2	0	2	1	1	1	1	1	1	**
1063	1063027	2	0	0	2	2	0	2	0	2	0	**
1063	1063028	2	0	0	2	1	1	0	2	1	1	**
1063	1063029	2	0	0	2	0	2	0	2	1	1	**
1063	1063031	0	2	0	2	2	0	1	1	2	0	**
1063	1063032	0	2	0	2	1	1	1	1	0	2	**
1063	1063033	2	0	1	1	2	0	2	0	2	0	**
1063	1063034	0	2	0	2	2	0	1	1	2	0	**
1063	1063035	1	1	1	1	2	0	2	0	1	1	**
1063	1063036	1	1	0	2	2	0	2	0	2	0	**
1063	1063037	1	1	0	2	2	0	0	2	1	1	**
1064	1064001	2	0	0	2	2	0	1	1	1	1	**
1064	1064002	2	0	0	2	2	0	1	1	1	1	**
1064	1064003	0	2	0	2	0	2	0	2	1	1	**
1064	1064004	1	1	0	2	2	0	1	1	1	1	**
1064	1064005	1	1	0	2	2	0	2	0	1	1	**
1064	1064006	2	0	0	2	2	0	2	0	2	0	**
1064	1064007	0	2	1	1	0	2	1	1	1	1	**
1064	1064008	0	2	0	2	2	0	2	0	*	*	**
1064	1064009	1	1	0	2	2	0	2	0	2	0	**
1064	1064010	2	0	0	2	1	1	1	1	1	1	**
1064	1064011	2	0	0	2	2	0	1	1	2	0	**
1064	1064012	1	1	0	2	1	1	2	0	2	0	**
1064	1064013	2	0	0	2	2	0	2	0	2	0	**
1064	1064014	0	2	1	1	1	1	0	2	0	2	**
1064	1064015	1	1	0	2	2	0	2	0	2	0	**
1064	1064016	2	0	0	2	2	0	2	0	2	0	**
1064	1064020	1	1	1	1	2	0	1	1	1	1	**
1064	1064022	2	0	0	2	2	0	2	0	1	1	**
1064	1064023	2	0	0	2	1	1	0	2	1	1	**
1064	1064024	2	0	1	1	2	0	2	0	0	2	**
1064	1064025	0	2	0	2	0	2	0	2	1	1	**
1064	1064026	2	0	1	1	2	0	2	0	2	0	**
1064	1064027	1	1	1	1	2	0	2	0	0	2	**
1064	1064028	2	0	0	2	2	0	2	0	2	0	**
1064	1064029	2	0	0	2	2	0	2	0	1	1	**
1064	1064031	2	0	1	1	2	0	2	0	2	0	**
1064	1064032	1	1	1	1	2	0	2	0	2	0	**
1064	1064033	2	0	1	1	2	0	2	0	1	1	**
1064	1064034	2	0	0	2	2	0	2	0	2	0	**
1064	1064037	2	0	0	2	2	0	2	0	2	0	**
1064	1064038	1	1	0	2	1	1	2	0	2	0	**
1064	1064040	1	1	0	2	2	0	1	1	2	0	**
1064	1064041	2	0	0	2	2	0	2	0	2	0	**
1066	1066001	2	0	0	2	2	0	2	0	0	2	**
1066	1066007	2	0	*	*	2	0	2	0	0	2	**
1066	1066008	2	0	*	*	2	0	2	0	0	2	**
1066	1066009	2	0	*	*	2	0	2	0	2	0	**
1066	1066010	2	0	*	*	2	0	2	0	2	0	**

Site	IND	AK b	AK v	GPI b	GPI v	MDH b	MDH v	LDH b	LDH v	IDH b	IDH v	Ck bv
1066	1066011	2	0	*	*	2	0	1	1	2	0	**
1066	1066012	0	2	*	*	2	0	1	1	2	0	**
1066	1066013	1	1	*	*	1	1	1	1	1	1	**
1067	1067001	0	2	0	2	1	1	0	2	0	2	**
1067	1067002	1	1	0	2	2	0	2	0	2	0	**
1067	1067003	2	0	0	2	0	2	0	2	0	2	**
1068	1068001	2	0	0	2	1	1	0	2	0	2	**
1069	1069001	1	1	*	*	2	0	1	1	2	0	**
1069	1069002	2	0	*	*	2	0	1	1	2	0	**
1069	1069003	2	0	*	*	2	0	0	2	2	0	**
1069	1069004	1	1	*	*	2	0	2	0	2	0	**
1069	1069005	2	0	*	*	2	0	2	0	2	0	**
1070	1070001	0	2	0	2	0	2	1	1	0	2	**
1070	1070002	0	2	0	2	1	1	1	1	1	1	**
1070	1070003	0	2	0	2	0	2	0	2	0	2	**
1070	1070004	0	2	0	2	0	2	0	2	1	1	**
1070	1070005	1	1	0	2	1	1	1	1	2	0	**
1070	1070006	0	2	0	2	0	2	0	2	0	2	**
1070	1070007	1	1	0	2	0	2	0	2	0	2	**
1070	1070008	1	1	0	2	0	2	0	2	0	2	**
1070	1070009	0	2	0	2	0	2	1	1	0	2	**
1071	1071001	0	2	1	1	0	2	0	2	0	2	**
1071	1071002	2	0	0	2	0	2	0	2	1	1	**
1071	1071003	0	2	0	2	0	2	0	2	0	2	**
1071	1071004	0	2	0	2	1	1	0	2	0	2	**
1071	1071005	0	2	0	2	0	2	0	2	0	2	**
1072	1072001	*	*	*	*	*	*	*	*	*	*	**
1072	1072002	*	*	*	*	*	*	*	*	*	*	**
1072	1072003	*	*	*	*	*	*	*	*	*	*	**
1072	1072004	*	*	*	*	*	*	*	*	*	*	**
1072	1072005	0	2	0	2	0	2	0	2	0	2	**
1072	1072006	0	2	0	2	0	2	0	2	1	1	**
1072	1072007	0	2	0	2	1	1	0	2	1	1	**
1073	1073001	1	1	0	2	0	2	0	2	0	2	**
1074	1074001	1	1	0	2	1	1	0	2	0	2	**
1074	1074002	0	2	0	2	0	2	1	1	0	2	**
1074	1074003	0	2	0	2	1	1	0	2	0	2	**
1074	1074004	1	1	0	2	0	2	0	2	0	2	**
1074	1074008	0	2	0	2	0	2	0	2	0	2	**
1075	1075001	2	0	0	2	0	2	0	2	0	2	**
1075	1075002	0	2	0	2	0	2	0	2	0	2	**
1076	1076001	0	2	0	2	0	2	0	2	0	2	**
1076	1076002	0	2	0	2	1	1	0	2	0	2	**
1077	1077001	1	1	2	0	0	2	0	2	*	*	**
1077	1077002	0	2	1	1	0	2	0	2	*	*	**
1077	1077003	0	2	1	1	0	2	1	1	*	*	**
1078	1078001	1	1	0	2	1	1	1	1	0	2	**
1078	1078002	0	2	0	2	1	1	1	1	0	2	**
1078	1078003	1	1	0	2	0	2	0	2	0	2	**
1078	1078004	0	2	0	2	0	2	1	1	0	2	**
1079	1079001	2	0	2	0	0	2	0	2	*	*	**
1079	1079002	0	2	0	2	0	2	0	2	*	*	**
1080	1080001	1	1	0	2	0	2	1	1	0	2	**
1081	1081001	2	0	*	*	2	0	1	1	2	0	**
1081	1081002	1	1	*	*	1	1	2	0	2	0	**
1081	1081003	2	0	*	*	2	0	2	0	1	1	**
1081	1081004	2	0	*	*	2	0	2	0	2	0	**
1082	1082001	2	0	*	*	2	0	0	2	1	1	**
1082	1082002	0	2	*	*	0	2	1	1	2	0	**
1083	1083001	0	2	0	2	0	2	1	1	2	0	**
1084	1084001	0	2	*	*	1	1	1	1	1	1	**
1084	1084002	2	0	*	*	2	0	1	1	2	0	**
1084	1084003	2	0	*	*	2	0	2	0	2	0	**
1085	1085001	1	1	0	2	1	1	0	2	2	0	**
1085	1085002	1	1	0	2	1	1	2	0	2	0	**
1085	1085003	1	1	0	2	1	1	0	2	2	0	**
1086	1086001	1	1	*	*	0	2	1	1	0	2	**
1087	1087001	0	2	*	*	0	2	1	1	1	1	**
1087	1087002	0	2	*	*	0	2	1	1	1	1	**
1087	1087003	0	2	*	*	0	2	0	2	0	2	**
1087	1087004	1	1	*	*	0	2	1	1	0	2	**
1087	1087005	0	2	*	*	1	1	0	2	0	2	**
1089	1089001	0	2	*	*	0	2	0	2	0	2	**
1091	1091001	0	2	0	2	0	2	0	2	0	2	**
1091	1091002	0	2	0	2	0	2	0	2	0	2	**
1091	1091003	1	1	0	2	0	2	0	2	0	2	**
1091	1091004	0	2	0	2	0	2	1	1	0	2	**
1092	1092001	0	2	*	*	1	1	0	2	0	2	**
1097	1097001	0	2	0	2	0	2	0	2	0	2	**
1097	1097002	0	2	0	2	0	2	0	2	0	2	**
1097	1097003	0	2	0	2	0	2	0	2	0	2	**
1097	1097004	0	2	0	2	0	2	1	1	2	0	**
1097	1097005	0	2	0	2	0	2	0	2	2	0	**
1097	1097006	0	2	0	2	0	2	0	2	2	0	**
1098	1098001	0	2	*	*	0	2	0	2	0	2	**
1099	1099001	0	2	0	2	1	1	0	2	0	2	**
1099	1099002	0	2	0	2	0	2	0	2	0	2	**
1099	1099003	0	2	0	2	0	2	1	1	0	2	**
1099	1099004	2	0	0	2	2	0	1	1	2	0	**
1099	1099005	0	2	0	2	1	1	0	2	0	2	**
1099	1099006	2	0	0	2	2	0	2	0	1	1	**
1099	1099007	0	2	0	2	1	1	0	2	0	2	**
1099	1099008	0	2	0	2	1	1	1	1	0	2	**
1099	1099011	0	2	0	2	0	2	1	1	0	2	**
1099	1099012	2	0	0	2	1	1	0	2	0	2	**
1099	1099014	0	2	0	2	0	2	0	2	0	2	**
1099	1099015	0	2	0	2	1	1	1	1	0	2	**
1099	1099017	2	0	0	2	0	2	0	2	0	2	**
1099	1099019	1	1	0	2	0	2	0	2	0	2	**
1099	1099021	0	2	0	2	1	1	0	2	0	2	**
1099	1099022	0	2	1	1	2	0	0	2	0	2	**
1099	1099023	0	2	0	2	0	2	0	2	0	2	**
1099	1099024	0	2	0	2	1	1	0	2	0	2	**
1099	1099027	0	2	0	2	1	1	0	2	0	2	**
1099	1099030	1	1	0	2	1	1	0	2	0	2	**
1099	1099036	0	2	0	2	1	1	0	2	0	2	**
1099	1099037	0	2	0	2	0	2	1	1	0	2	**
1104	1104021	2	0	0	2	2	0	1	1	2	0	**
1104	1104022	2	0	0	2	2	0	2	0	2	0	**
1104	1104023	2	0	1	1	2	0	2	0	2	0	**
1104	1104026	2	0	0	2	2	0	2	0	2	0	**
1104	1104029	2	0	0	2	2	0	1	1	1	1	**
1104	1104030	0	2	0	2	0	2	0	2	0	2	**
1104	1104031	2	0	0	2	2	0	2	0	2	0	**
1104	1104033	2	0	0	2	2	0	2	0	2	0	**
1104	1104034	2	0	1	1	2	0	2	0	2	0	**
1104	1104035	1	1	0	2	2	0	2	0	2	0	**
1104	1104037	1	1	0	2	2	0	1	1	2	0	**
1104	1104039	0	2	1	1	0	2	0	2	1	1	**

Site	IND	AK b	AK v	GPI b	GPI v	MDH b	MDH v	LDH b	LDH v	IDH b	IDH v	Ck bv
1104	1104040	2	0	0	2	2	0	2	0	2	0	**
1105	1105009	2	0	0	2	1	1	0	2	1	1	**
1105	1105010	2	0	1	1	2	0	2	0	2	0	**
1105	1105011	2	0	0	2	2	0	2	0	2	0	**
1109	1109003	2	0	1	1	2	0	2	0	2	0	**
1110	1110001	1	1	0	2	2	0	0	2	0	2	**
1110	1110002	2	0	0	2	1	1	0	2	0	2	**
1110	1110003	0	2	0	2	0	2	0	2	1	1	**
1110	1110006	2	0	0	2	2	0	2	0	2	0	**
1110	1110007	1	1	0	2	0	2	0	2	0	2	**
1110	1110008	1	1	0	2	1	1	1	1	0	2	**
1110	1110009	0	2	0	2	0	2	0	2	1	1	**
1110	1110011	0	2	0	2	0	2	0	2	0	2	**
1110	1110012	2	0	0	2	2	0	0	2	2	0	**
1110	1110014	1	1	0	2	1	1	0	2	0	2	**
1110	1110015	1	1	0	2	2	0	1	1	1	1	**
1110	1110016	1	1	0	2	1	1	1	1	1	1	**
1110	1110017	2	0	0	2	2	0	2	0	1	1	**
1111	1111002	0	2	0	2	0	2	0	2	0	2	**
1111	1111004	0	2	0	2	0	2	0	2	0	2	**
1112	1112001	0	2	0	2	0	2	0	2	0	2	**
1112	1112002	0	2	0	2	0	2	0	2	0	2	**
1112	1112005	1	1	0	2	2	0	0	2	0	2	**
1112	1112009	0	2	0	2	1	1	0	2	0	2	**
1112	1112010	0	2	0	2	0	2	0	2	0	2	**
1112	1112012	1	1	0	2	2	0	0	2	0	2	**
1113	1113002	0	2	0	2	1	1	1	1	1	1	**
1113	1113003	0	2	0	2	1	1	0	2	0	2	**
1113	1113004	1	1	0	2	1	1	2	0	1	1	**
1113	1113005	0	2	0	2	1	1	1	1	1	1	**
1113	1113006	1	1	0	2	2	0	1	1	2	0	**
1113	1113007	0	2	1	1	0	2	0	2	*	*	**
2001	2001001	2	0	0	2	0	2	1	1	0	2	**
2001	2001002	1	1	0	2	1	1	0	2	0	2	**
2001	2001003	1	1	0	2	0	2	0	2	1	1	**
2001	2001004	0	2	0	2	1	1	2	0	*	*	**
2001	2001005	1	1	0	2	0	2	1	1	*	*	**
2001	2001006	0	2	1	1	1	1	0	2	*	*	**
2001	2001007	2	0	0	2	0	2	1	1	*	*	**
2002	2002001	1	1	0	2	0	2	1	1	1	1	**
2002	2002002	1	1	0	2	1	1	0	2	0	2	**
2002	2002003	2	0	1	1	2	0	1	1	1	1	**
2002	2002004	2	0	0	2	0	2	2	0	0	2	**
2002	2002005	1	1	1	1	1	1	1	1	0	2	**
2002	2002006	2	0	0	2	0	2	0	2	0	2	**
2002	2002007	2	0	0	2	0	2	0	2	1	1	**
2002	2002008	1	1	0	2	1	1	0	2	0	2	**
2002	2002009	1	1	0	2	0	2	0	2	*	*	**
2002	2002010	1	1	1	1	1	1	1	1	1	1	**
2002	2002011	0	2	0	2	1	1	1	1	0	2	**
2002	2002012	1	1	0	2	0	2	2	0	*	*	**
2002	2002013	1	1	0	2	0	2	0	2	*	*	**
2002	2002014	1	1	0	2	0	2	1	1	*	*	**
2002	2002015	1	1	0	2	0	2	0	2	*	*	**
2002	2002016	0	2	0	2	0	2	0	2	*	*	**
2002	2002017	1	1	0	2	0	2	0	2	*	*	**
2002	2002018	2	0	0	2	1	1	1	1	0	2	**
2003	2003001	1	1	0	2	0	2	0	2	0	2	**
2003	2003002	1	1	0	2	0	2	1	1	1	1	**
2003	2003003	0	2	0	2	0	2	0	2	0	2	**
2003	2003004	0	2	0	2	0	2	1	1	1	1	**
2003	2003005	0	2	0	2	0	2	1	1	1	1	**
2003	2003006	1	1	0	2	1	1	0	2	1	1	**
2003	2003007	1	1	0	2	0	2	1	1	0	2	**
2003	2003008	1	1	0	2	1	1	1	1	1	1	**
2003	2003009	0	2	0	2	1	1	0	2	0	2	**
2003	2003010	0	2	0	2	2	0	0	2	0	2	**
2003	2003011	1	1	0	2	0	2	1	1	0	2	**
2003	2003012	1	1	0	2	0	2	0	2	0	2	**
2003	2003013	1	1	0	2	0	2	0	2	1	1	**
2003	2003014	0	2	0	2	0	2	0	2	*	*	**
2003	2003015	1	1	0	2	0	2	0	2	*	*	**
2003	2003016	1	1	0	2	1	1	0	2	*	*	**
2003	2003017	*	*	0	2	0	2	0	2	0	2	**
2003	2003018	*	*	0	2	1	1	1	1	0	2	**
2003	2003019	1	1	0	2	2	0	0	2	*	*	**
2003	2003020	1	1	1	1	0	2	1	1	0	2	**
2003	2003021	1	1	0	2	0	2	1	1	0	2	**
2003	2003022	2	0	0	2	1	1	1	1	0	2	**
2003	2003023	1	1	0	2	0	2	0	2	0	2	**
2003	2003024	1	1	0	2	0	2	0	2	*	*	**
2011	2011001	2	0	0	2	0	2	1	1	1	1	**
2011	2011002	2	0	0	2	2	0	2	0	2	0	**
2011	2011003	1	1	1	1	1	1	1	1	1	1	**
2011	2011004	1	1	0	2	0	2	0	2	1	1	**
2011	2011005	2	0	0	2	2	0	2	0	2	0	**
2011	2011006	2	0	0	2	2	0	1	1	2	0	**
2011	2011007	0	2	0	2	0	2	0	2	0	2	**
2011	2011008	2	0	0	2	2	0	2	0	2	0	**
2011	2011009	2	0	0	2	2	0	2	0	2	0	**
2011	2011010	0	2	0	2	1	1	0	2	2	0	**
2011	2011011	1	1	0	2	0	2	1	1	0	2	**
2011	2011012	1	1	0	2	0	2	1	1	1	1	**
2011	2011013	2	0	0	2	1	1	1	1	1	1	**
2011	2011014	2	0	0	2	2	0	2	0	2	0	**
2011	2011016	2	0	0	2	0	2	1	1	0	2	**
2011	2011017	2	0	0	2	2	0	2	0	2	0	**
2011	2011019	1	1	0	2	0	2	1	1	1	1	**
2011	2011020	1	1	0	2	0	2	1	1	0	2	**
2011	2011021	0	2	0	2	0	2	1	1	1	1	**
2011	2011022	2	0	1	1	2	0	2	0	2	0	**
2011	2011024	2	0	1	1	2	0	2	0	2	0	**
2012	2012001	*	*	1	1	2	0	2	0	2	0	**
2012	2012002	*	*	0	2	1	1	1	1	2	0	**
2012	2012003	1	1	0	2	1	1	1	1	1	1	**
2012	2012004	0	2	0	2	0	2	1	1	0	2	**
2013	2013001	0	2	0	2	0	2	1	1	0	2	**
2013	2013002	2	0	0	2	2	0	2	0	2	0	**
2013	2013003	1	1	0	2	1	1	1	1	1	1	**
2013	2013004	2	0	0	2	2	0	2	0	1	1	**
2033	2033001	2	0	0	2	2	0	2	0	*	*	**
2033	2033002	2	0	0	2	2	0	2	0	*	*	**
2033	2033003	2	0	0	2	2	0	2	0	*	*	**
2033	2033004	2	0	1	1	2	0	2	0	*	*	**
2033	2033005	2	0	0	2	2	0	1	1	2	0	**
2033	2033006	2	0	0	2	2	0	1	1	2	0	**
2033	2033007	2	0	1	1	2	0	2	0	2	0	**
2033	2033008	0	2	0	2	2	0	1	1	2	0	**
2033	2033009	2	0	0	2	2	0	2	0	2	0	**

Site	IND	AK b	AK v	GPI b	GPI v	MDH b	MDH v	LDH b	LDH v	IDH b	IDH v	Ck bv
2033	2033010	2	0	0	2	2	0	2	0	2	0	**
2033	2033011	2	0	1	1	2	0	2	0	2	0	**
2033	2033012	2	0	0	2	2	0	1	1	2	0	**
2033	2033013	2	0	0	2	2	0	1	1	1	1	**
2039	2039001	2	0	2	0	2	0	2	0	2	0	**
2039	2039002	1	1	0	2	2	0	2	0	2	0	**
2039	2039003	2	0	0	2	2	0	2	0	2	0	**
2039	2039004	2	0	0	2	2	0	2	0	*	*	**
2039	2039007	2	0	0	2	2	0	2	0	2	0	**
2040	2040001	2	0	0	2	2	0	2	0	2	0	**
2040	2040002	2	0	0	2	2	0	2	0	2	0	**
2040	2040003	1	1	0	2	2	0	2	0	2	0	**
2040	2040004	2	0	1	1	2	0	2	0	2	0	**
2043	2043001	2	0	0	2	2	0	2	0	0	2	**
2043	2043002	1	1	1	1	2	0	2	0	1	1	**
2043	2043003	2	0	0	2	2	0	2	0	2	0	**
2054	2054001	2	0	0	2	1	1	2	0	1	1	**
2054	2054002	2	0	0	2	2	0	2	0	1	1	**
2054	2054003	2	0	1	1	2	0	2	0	2	0	**
2055	2055001	1	1	0	2	2	0	2	0	2	0	**
2055	2055002	2	0	0	2	2	0	2	0	2	0	**
2055	2055003	2	0	0	2	2	0	2	0	2	0	**
2055	2055004	2	0	0	2	2	0	2	0	2	0	**
2055	2055005	2	0	0	2	2	0	2	0	2	0	**
2055	2055006	2	0	0	2	2	0	2	0	2	0	**
2055	2055007	2	0	0	2	2	0	1	1	2	0	**
2063	2063001	1	1	*	*	2	0	2	0	2	0	**
2074	2074001	1	1	0	2	0	2	0	2	*	*	**
2074	2074002	1	1	0	2	1	1	0	2	*	*	**
2074	2074003	2	0	0	2	0	2	0	2	*	*	**
2074	2074004	0	2	0	2	0	2	0	2	*	*	**
2074	2074005	0	2	0	2	1	1	0	2	*	*	**
2074	2074006	0	2	0	2	0	2	0	2	0	2	**
2074	2074007	1	1	0	2	0	2	0	2	0	2	**
2074	2074008	1	1	0	2	0	2	1	1	0	2	**
2074	2074009	0	2	0	2	0	2	0	2	0	2	**
2074	2074010	0	2	0	2	0	2	0	2	0	2	**
2074	2074011	0	2	0	2	0	2	0	2	2	0	**
2074	2074012	1	1	0	2	0	2	0	2	1	1	**
2074	2074013	0	2	0	2	1	1	1	1	0	2	**
2074	2074014	0	2	0	2	0	2	1	1	2	0	**
2074	2074015	1	1	0	2	0	2	0	2	0	2	**
2074	2074016	0	2	0	2	0	2	0	2	1	1	**
2074	2074017	0	2	0	2	0	2	0	2	0	2	**
2074	2074018	0	2	0	2	0	2	0	2	0	2	**
2074	2074019	0	2	0	2	1	1	0	2	1	1	**
2074	2074020	0	2	0	2	0	2	0	2	0	2	**
2074	2074021	1	1	0	2	0	2	1	1	0	2	**
2074	2074022	1	1	0	2	1	1	0	2	0	2	**
2074	2074023	0	2	1	1	0	2	0	2	1	1	**
2074	2074024	0	2	0	2	0	2	0	2	1	1	**
2074	2074025	0	2	0	2	0	2	0	2	1	1	**
2074	2074026	0	2	0	2	0	2	0	2	0	2	**
2074	2074027	0	2	0	2	0	2	0	2	0	2	**
2074	2074028	1	1	0	2	0	2	0	2	1	1	**
2074	2074029	1	1	0	2	0	2	0	2	0	2	**
2074	2074030	0	2	0	2	0	2	0	2	0	2	**
2074	2074031	1	1	0	2	1	1	0	2	0	2	**
2074	2074032	0	2	0	2	0	2	0	2	0	2	**
2074	2074033	2	0	0	2	2	0	1	1	0	2	**
2074	2074034	0	2	0	2	0	2	0	2	0	2	**
2074	2074035	0	2	0	2	0	2	0	2	0	2	**
2074	2074036	1	1	0	2	1	1	0	2	0	2	**
2074	2074037	1	1	0	2	1	1	0	2	0	2	**
2074	2074038	1	1	0	2	0	2	0	2	0	2	**
2074	2074039	0	2	0	2	1	1	1	1	0	2	**
2074	2074040	0	2	0	2	1	1	1	1	0	2	**
2074	2074041	0	2	0	2	0	2	0	2	*	*	**
2074	2074042	1	1	0	2	0	2	0	2	*	*	**
2074	2074043	1	1	0	2	0	2	0	2	*	*	**
2074	2074044	1	1	0	2	0	2	0	2	*	*	**
2074	2074045	1	1	0	2	0	2	0	2	*	*	**
2082	2082001	2	0	0	2	2	0	2	0	2	0	**
2082	2082002	2	0	0	2	2	0	2	0	2	0	**
2099	2099001	1	1	0	2	0	2	0	2	1	1	**
2099	2099002	0	2	0	2	0	2	0	2	0	2	**
2099	2099003	0	2	0	2	0	2	0	2	0	2	**
2099	2099004	0	2	0	2	0	2	1	1	1	1	**
2099	2099005	1	1	0	2	1	1	0	2	1	1	**
2099	2099006	0	2	0	2	1	1	0	2	0	2	**
2099	2099007	1	1	0	2	0	2	1	1	0	2	**
2099	2099008	1	1	0	2	0	2	0	2	0	2	**
2099	2099009	1	1	0	2	1	1	1	1	0	2	**
2099	2099010	1	1	0	2	0	2	0	2	1	1	**
2100	2100001	1	1	0	2	0	2	1	1	0	2	**
2100	2100002	0	2	0	2	0	2	1	1	0	2	**
2100	2100003	0	2	0	2	0	2	0	2	0	2	**
2100	2100004	0	2	0	2	2	0	0	2	1	1	**
2100	2100005	0	2	0	2	0	2	0	2	0	2	**
2100	2100006	2	0	0	2	0	2	0	2	0	2	**
2103	2103001	2	0	0	2	2	0	2	0	2	0	**
2103	2103002	2	0	0	2	2	0	2	0	2	0	**
2103	2103003	2	0	0	2	1	1	0	2	0	2	**
2103	2103004	1	1	0	2	1	1	1	1	2	0	**
2103	2103005	2	0	0	2	2	0	2	0	2	0	**
2103	2103006	1	1	0	2	1	1	1	1	1	1	**
2103	2103008	2	0	0	2	2	0	2	0	2	0	**
2103	2103009	2	0	0	2	2	0	2	0	2	0	**
2103	2103010	1	1	0	2	1	1	0	2	0	2	**
2103	2103011	2	0	0	2	2	0	2	0	2	0	**
2103	2103023	2	0	0	2	2	0	1	1	1	1	**
2115	2115001	2	0	0	2	2	0	2	0	*	*	**
2115	2115002	1	1	0	2	2	0	1	1	2	0	**
2115	2115003	2	0	0	2	2	0	1	1	2	0	**
2115	2115004	2	0	1	1	2	0	2	0	1	1	**
2115	2115005	2	0	1	1	2	0	1	1	2	0	**
2115	2115006	2	0	0	2	2	0	2	0	2	0	**
2115	2115007	2	0	1	1	2	0	2	0	2	0	**
2116	2116001	1	1	1	1	2	0	1	1	2	0	**
2116	2116002	2	0	0	2	2	0	1	1	2	0	**
2116	2116003	2	0	0	2	2	0	2	0	2	0	**
2116	2116004	2	0	1	1	2	0	2	0	1	1	**
2116	2116005	2	0	1	1	2	0	2	0	2	0	**
2116	2116006	2	0	1	1	2	0	2	0	2	0	**
2116	2116007	1	1	0	2	2	0	1	1	2	0	**
2116	2116008	2	0	0	2	2	0	2	0	2	0	**
2116	2116009	2	0	0	2	2	0	1	1	2	0	**
2116	2116010	2	0	0	2	2	0	2	0	2	0	**
2116	2116011	2	0	1	1	2	0	2	0	*	*	**

Site	IND	AK b	AK v	GPI b	GPI v	MDH b	MDH v	LDH b	LDH v	IDH b	IDH v	Ck bv
2116	2116012	2	0	0	2	2	0	2	0	*	*	**
2116	2116013	2	0	0	2	2	0	1	1	*	*	**
2116	2116014	2	0	0	2	2	0	2	0	*	*	**
2116	2116015	2	0	0	2	2	0	2	0	2	0	**
2116	2116016	2	0	0	2	2	0	1	1	2	0	**
2116	2116017	1	1	1	1	2	0	2	0	2	0	**
2116	2116018	1	1	0	2	2	0	2	0	2	0	**
2117	2117001	1	1	0	2	2	0	2	0	2	0	**
2117	2117002	2	0	1	1	2	0	2	0	2	0	**
2117	2117003	1	1	0	2	2	0	2	0	2	0	**
2117	2117004	2	0	0	2	2	0	2	0	2	0	**
2117	2117005	2	0	1	1	2	0	2	0	2	0	**
2117	2117006	2	0	0	2	1	1	2	0	*	*	**
2117	2117007	2	0	0	2	2	0	1	1	*	*	**
2117	2117008	2	0	1	1	1	1	1	1	*	*	**
2117	2117009	2	0	1	1	2	0	2	0	*	*	**
2117	2117010	2	0	1	1	2	0	2	0	*	*	**
2117	2117012	2	0	0	2	2	0	2	0	2	0	**
2117	2117013	1	1	0	2	2	0	2	0	2	0	**
2117	2117014	2	0	0	2	2	0	2	0	2	0	**
2118	2118001	0	2	0	2	0	2	0	2	0	2	**
2118	2118002	2	0	0	2	1	1	0	2	0	2	**
2119	2119001	1	1	0	2	2	0	2	0	2	0	**
2119	2119002	2	0	0	2	2	0	1	1	1	1	**
2119	2119003	2	0	1	1	2	0	2	0	1	1	**
2119	2119004	2	0	0	2	2	0	2	0	2	0	**
2119	2119005	2	0	1	1	2	0	2	0	2	0	**
2119	2119006	2	0	0	2	2	0	2	0	2	0	**
2119	2119007	2	0	2	0	2	0	2	0	2	0	**
2120	2120001	2	0	0	2	2	0	1	1	2	0	**
2120	2120002	0	2	0	2	2	0	1	1	2	0	**
2120	2120003	1	1	0	2	2	0	2	0	2	0	**
2120	2120004	2	0	0	2	2	0	2	0	2	0	**
2120	2120005	2	0	0	2	2	0	2	0	*	*	**
2120	2120006	2	0	0	2	2	0	2	0	*	*	**
2121	2121001	2	0	0	2	2	0	2	0	*	*	**
2121	2121002	2	0	0	2	2	0	1	1	*	*	**
2121	2121003	2	0	0	2	2	0	2	0	*	*	**
2121	2121004	2	0	0	2	2	0	2	0	*	*	**
2121	2121005	2	0	0	2	2	0	2	0	*	*	**
2121	2121015	2	0	0	2	2	0	2	0	*	*	**
2121	2121016	2	0	0	2	2	0	2	0	*	*	**
2121	2121017	2	0	0	2	2	0	2	0	2	0	**
2122	2122009	0	2	0	2	0	2	0	2	0	2	**
2122	2122010	0	2	0	2	0	2	0	2	1	1	**
2122	2122011	0	2	0	2	0	2	0	2	1	1	**
2122	2122012	0	2	0	2	0	2	0	2	0	2	**
2124	2124011	1	1	0	2	0	2	0	2	1	1	**
2124	2124013	1	1	0	2	0	2	0	2	0	2	**
2126	2126011	2	0	0	2	0	2	0	2	0	2	**
2126	2126012	1	1	0	2	1	1	0	2	0	2	**
2126	2126013	1	1	0	2	0	2	0	2	0	2	**
2126	2126014	1	1	0	2	1	1	0	2	0	2	**
2126	2126015	1	1	0	2	0	2	0	2	0	2	**
2126	2126016	2	0	0	2	0	2	0	2	0	2	**
2126	2126017	1	1	0	2	0	2	0	2	0	2	**
2126	2126018	0	2	0	2	0	2	0	2	0	2	**
2127	2127001	0	2	0	2	0	2	0	2	1	1	**
2127	2127003	0	2	0	2	0	2	0	2	0	2	**
2127	2127010	0	2	0	2	0	2	0	2	0	2	**
2127	2127015	0	2	0	2	0	2	0	2	0	2	**
2132	2132005	1	1	0	2	0	2	0	2	0	2	**
2132	2132006	0	2	0	2	0	2	0	2	0	2	**
2133	2133001	0	2	0	2	0	2	0	2	0	2	**
2133	2133002	0	2	0	2	0	2	0	2	0	2	**
2133	2133003	0	2	0	2	1	1	1	1	0	2	**
2133	2133015	0	2	0	2	0	2	1	1	0	2	**
2133	2133019	1	1	0	2	0	2	0	2	0	2	**
2133	2133020	0	2	0	2	0	2	1	1	0	2	**
2133	2133021	0	2	0	2	0	2	0	2	0	2	**
2133	2133024	0	2	0	2	0	2	0	2	0	2	**
2134	2134001	1	1	0	2	0	2	0	2	0	2	**
2134	2134002	0	2	0	2	0	2	0	2	0	2	**
2134	2134003	1	1	0	2	0	2	0	2	0	2	**
2134	2134004	1	1	0	2	0	2	1	1	1	1	**
2134	2134005	0	2	0	2	0	2	0	2	0	2	**
2134	2134013	1	1	0	2	0	2	0	2	*	*	**
2134	2134014	0	2	0	2	0	2	0	2	1	1	**
2134	2134015	1	1	0	2	0	2	0	2	0	2	**
2134	2134017	0	2	0	2	0	2	0	2	0	2	**
2135	2135001	1	1	0	2	2	0	2	0	2	0	**
2135	2135002	2	0	0	2	2	0	2	0	2	0	**
2135	2135003	2	0	0	2	2	0	2	0	2	0	**
2135	2135004	1	1	0	2	2	0	1	1	1	1	**
2135	2135005	2	0	0	2	2	0	1	1	*	*	**
2135	2135006	2	0	0	2	2	0	2	0	*	*	**
2135	2135007	2	0	0	2	2	0	1	1	*	*	**
2135	2135008	2	0	0	2	2	0	2	0	*	*	**
2135	2135009	2	0	1	1	2	0	2	0	*	*	**
2135	2135010	2	0	0	2	1	1	2	0	*	*	**
2135	2135011	2	0	1	1	2	0	2	0	*	*	**
2136	2136001	0	2	0	2	0	2	0	2	*	*	**
2136	2136002	1	1	0	2	0	2	0	2	0	2	**
2136	2136003	1	1	0	2	0	2	0	2	*	*	**
2138	2138007	1	1	0	2	0	2	0	2	0	2	**
2140	2140001	1	1	0	2	0	2	0	2	1	1	**
2140	2140002	1	1	0	2	0	2	0	2	1	1	**
2140	2140003	1	1	0	2	0	2	0	2	0	2	**
2140	2140004	1	1	0	2	0	2	0	2	0	2	**
2140	2140006	2	0	0	2	0	2	1	1	*	*	**
2140	2140007	1	1	0	2	0	2	0	2	*	*	**
2141	2141003	1	1	0	2	0	2	0	2	*	*	**
2141	2141004	0	2	0	2	0	2	0	2	*	*	**
2141	2141005	1	1	0	2	0	2	0	2	*	*	**
2141	2141006	1	1	0	2	1	1	0	2	0	2	**
2142	2142001	2	0	0	2	2	0	2	0	2	0	**
2142	2142002	1	1	0	2	1	1	1	1	1	1	**
2142	2142003	2	0	0	2	2	0	2	0	2	0	**
2143	2143001	2	0	0	2	1	1	1	1	*	*	**
2143	2143002	2	0	0	2	2	0	2	0	*	*	**
2143	2143003	2	0	0	2	2	0	2	0	*	*	**
2143	2143004	1	1	1	1	2	0	2	0	*	*	**
2143	2143005	2	0	0	2	2	0	2	0	*	*	**
2143	2143006	1	1	0	2	2	0	2	0	*	*	**
2143	2143007	2	0	1	1	2	0	1	1	*	*	**
2143	2143008	2	0	0	2	2	0	2	0	*	*	**
2143	2143009	1	1	0	2	2	0	0	2	*	*	**
2143	2143010	2	0	1	1	2	0	2	0	*	*	**
2143	2143011	2	0	0	2	2	0	2	0	*	*	**

Site	IND	AK b	AK v	GPI b	GPI v	MDH b	MDH v	LDH b	LDH v	IDH b	IDH v	Ck bv
2143	2143012	1	1	1	1	2	0	2	0	*	*	**
2143	2143013	2	0	1	1	2	0	2	0	*	*	**
2143	2143014	1	1	0	2	2	0	2	0	*	*	**
2143	2143015	2	0	1	1	2	0	2	0	*	*	**
2143	2143016	2	0	0	2	2	0	2	0	*	*	**
2143	2143017	2	0	0	2	2	0	2	0	*	*	**
2143	2143018	2	0	0	2	2	0	2	0	*	*	**
2143	2143019	2	0	0	2	2	0	1	1	*	*	**
2143	2143020	2	0	0	2	2	0	2	0	2	0	**
2143	2143021	2	0	0	2	2	0	2	0	2	0	**
2143	2143022	1	1	0	2	2	0	2	0	2	0	**
2143	2143023	2	0	0	2	2	0	1	1	2	0	**
2143	2143024	2	0	0	2	2	0	2	0	2	0	**
2143	2143025	2	0	0	2	2	0	2	0	2	0	**
2143	2143026	1	1	0	2	2	0	2	0	2	0	**
2143	2143027	2	0	0	2	2	0	2	0	2	0	**
2143	2143028	2	0	0	2	2	0	1	1	2	0	**
2143	2143029	2	0	1	1	2	0	2	0	2	0	**
2143	2143030	*	*	1	1	*	*	1	1	*	*	**
2143	2143031	*	*	0	2	*	*	2	0	*	*	**
2143	2143032	0	2	1	1	0	2	1	1	0	2	**
2143	2143033	*	*	1	1	*	*	2	0	*	*	**
2143	2143034	1	1	0	2	2	0	2	0	*	*	**
2143	2143035	2	0	0	2	2	0	2	0	*	*	**
2143	2143036	2	0	0	2	2	0	2	0	*	*	**
2143	2143037	1	1	0	2	1	1	1	1	*	*	**
2143	2143038	1	1	0	2	2	0	2	0	*	*	**
2143	2143039	1	1	0	2	1	1	0	2	1	1	**
2144	2144001	*	*	0	2	*	*	0	2	*	*	**
2144	2144002	*	*	0	2	*	*	0	2	*	*	**
2145	2145001	2	0	0	2	2	0	2	0	2	0	**
2145	2145002	1	1	0	2	1	1	1	1	2	0	**
2145	2145003	2	0	0	2	2	0	2	0	2	0	**
2145	2145004	2	0	0	2	2	0	2	0	1	1	**
2145	2145005	1	1	0	2	0	2	0	2	1	1	**
2145	2145006	2	0	1	1	2	0	2	0	2	0	**
2146	2146001	2	0	0	2	0	2	0	2	0	2	**
2146	2146002	1	1	2	0	0	2	0	2	0	2	**
2146	2146003	0	2	0	2	0	2	0	2	2	0	**
2146	2146004	0	2	0	2	1	1	0	2	1	1	**
2146	2146006	0	2	0	2	0	2	0	2	0	2	**
2146	2146007	1	1	1	1	0	2	1	1	1	1	**
2147	2147001	2	0	0	2	2	0	2	0	2	0	**
2147	2147002	2	0	1	1	2	0	2	0	1	1	**
2147	2147003	2	0	0	2	0	2	0	2	1	1	**
2147	2147004	0	2	0	2	0	2	0	2	1	1	**
2147	2147005	2	0	1	1	1	1	0	2	0	2	**
2147	2147006	1	1	0	2	1	1	0	2	1	1	**
2147	2147007	0	2	0	2	1	1	0	2	0	2	**
2147	2147008	2	0	0	2	2	0	2	0	2	0	**
2147	2147009	1	1	0	2	0	2	0	2	*	*	**
2147	2147010	2	0	1	1	2	0	2	0	2	0	**
2148	2148001	2	0	0	2	2	0	2	0	2	0	**
2148	2148002	1	1	0	2	2	0	1	1	2	0	**
2148	2148003	2	0	0	2	2	0	2	0	1	1	**
2148	2148004	2	0	0	2	2	0	2	0	1	1	**
2149	2149001	1	1	0	2	0	2	1	1	2	0	**
2150	2150001	*	*	0	2	2	0	2	0	2	0	**
2150	2150002	*	*	0	2	1	1	1	1	1	1	**
2150	2150003	2	0	0	2	2	0	2	0	0	2	**
2150	2150004	2	0	1	1	2	0	2	0	2	0	**
2150	2150005	1	1	0	2	2	0	1	1	1	1	**
2151	2151001	2	0	0	2	2	0	2	0	2	0	**
2151	2151002	2	0	1	1	2	0	2	0	2	0	**
2151	2151003	2	0	1	1	2	0	2	0	2	0	**
2151	2151004	2	0	0	2	2	0	2	0	2	0	**
2151	2151005	2	0	0	2	2	0	2	0	2	0	**
2151	2151006	2	0	1	1	2	0	1	1	2	0	**
2151	2151007	2	0	0	2	2	0	0	2	2	0	**
2151	2151008	2	0	0	2	2	0	0	2	2	0	**
2151	2151009	2	0	0	2	2	0	1	1	1	1	**
2152	2152001	2	0	0	2	2	0	2	0	2	0	**
2152	2152002	1	1	0	2	2	0	1	1	1	1	**
2152	2152003	2	0	0	2	2	0	2	0	2	0	**
2152	2152004	2	0	0	2	2	0	2	0	2	0	**
2152	2152005	2	0	0	2	2	0	1	1	2	0	**
2152	2152006	2	0	0	2	2	0	2	0	2	0	**
2152	2152007	1	1	0	2	2	0	2	0	2	0	**
2152	2152008	2	0	0	2	2	0	2	0	2	0	**
2152	2152009	2	0	0	2	2	0	1	1	2	0	**
2152	2152010	2	0	1	1	2	0	1	1	2	0	**
2152	2152011	2	0	1	1	2	0	2	0	1	1	**
2152	2152012	2	0	0	2	2	0	1	1	2	0	**
2152	2152013	2	0	1	1	2	0	1	1	2	0	**
2152	2152014	1	1	1	1	1	1	*	*	1	1	**
2152	2152015	2	0	0	2	2	0	2	0	2	0	**
2152	2152016	1	1	1	1	2	0	0	2	2	0	**
2152	2152017	1	1	0	2	2	0	2	0	2	0	**
2152	2152018	2	0	0	2	2	0	2	0	*	*	**
2152	2152019	1	1	0	2	2	0	2	0	*	*	**
2152	2152020	2	0	0	2	2	0	2	0	*	*	**
2152	2152021	2	0	0	2	2	0	2	0	*	*	**
2152	2152022	2	0	0	2	2	0	2	0	*	*	**
2152	2152023	2	0	0	2	2	0	2	0	2	0	**
2152	2152024	2	0	0	2	2	0	1	1	2	0	**
2152	2152025	2	0	0	2	2	0	2	0	2	0	**
2152	2152026	2	0	0	2	2	0	2	0	2	0	**
2152	2152027	2	0	0	2	2	0	2	0	2	0	**
2152	2152028	2	0	0	2	2	0	2	0	1	1	**
2153	2153001	2	0	0	2	0	2	2	0	2	0	**
2153	2153002	1	1	0	2	0	2	2	0	2	0	**
2153	2153003	2	0	0	2	2	0	1	1	2	0	**
2154	2154001	1	1	0	2	1	1	0	2	1	1	**
2154	2154002	0	2	0	2	0	2	0	2	0	2	**
2154	2154003	1	1	0	2	1	1	0	2	0	2	**
2154	2154004	1	1	0	2	1	1	2	0	1	1	**
2154	2154005	0	2	0	2	1	1	0	2	*	*	**
2154	2154006	1	1	0	2	0	2	0	2	*	*	**
2154	2154007	0	2	0	2	0	2	1	1	*	*	**
2154	2154008	2	0	0	2	2	0	1	1	*	*	**
2154	2154009	0	2	0	2	0	2	0	2	*	*	**
2154	2154010	0	2	0	2	0	2	0	2	*	*	**
2154	2154011	0	2	0	2	1	1	0	2	*	*	**
2154	2154012	1	1	0	2	0	2	0	2	*	*	**
2154	2154013	0	2	1	1	1	1	0	2	*	*	**
2154	2154014	1	1	0	2	0	2	0	2	*	*	**
2154	2154015	1	1	0	2	0	2	1	1	*	*	**
2155	2155001	1	1	0	2	0	2	0	2	0	2	**
2155	2155002	1	1	0	2	1	1	0	2	1	1	**
2155	2155003	1	1	0	2	0	2	0	2	0	2	**

		AK		GPI		MDH		LDH		IDH		Ck
Site	IND	b	v	b	v	b	v	b	v	b	v	bv
2155	2155004	1	1	0	2	0	2	0	2	0	2	**
2156	2156001	1	1	0	2	0	2	1	1	0	2	**
2156	2156002	1	1	0	2	0	2	0	2	0	2	**
2156	2156003	0	2	0	2	0	2	0	2	0	2	**
2156	2156004	0	2	0	2	0	2	0	2	0	2	**
2156	2156005	1	1	0	2	2	0	0	2	0	2	**
2156	2156006	0	2	0	2	0	2	0	2	0	2	**
2156	2156007	0	2	0	2	0	2	0	2	0	2	**
2156	2156008	1	1	0	2	0	2	0	2	1	1	**
2156	2156009	2	0	0	2	0	2	0	2	0	2	**
2156	2156010	1	1	0	2	0	2	0	2	0	2	**
2156	2156011	0	2	0	2	1	1	0	2	0	2	**
2156	2156012	0	2	0	2	0	2	1	1	1	1	**
2157	2157001	2	0	0	2	2	0	2	0	2	0	**
2157	2157002	1	1	1	1	2	0	2	0	2	0	**
2157	2157003	2	0	0	2	2	0	0	2	2	0	**
2158	2158001	0	2	0	2	0	2	0	2	0	2	**
2158	2158002	0	2	0	2	0	2	0	2	0	2	**
2158	2158003	1	1	0	2	0	2	0	2	0	2	**
2158	2158004	1	1	0	2	0	2	0	2	0	2	**
2158	2158005	2	0	0	2	1	1	0	2	1	1	**
2159	2159001	2	0	1	1	2	0	2	0	2	0	**
2159	2159002	2	0	0	2	2	0	2	0	2	0	**
2159	2159003	2	0	0	2	2	0	2	0	2	0	**
2159	2159004	0	2	2	0	2	0	2	0	2	0	**
2159	2159005	2	0	0	2	2	0	2	0	2	0	**
2159	2159006	2	0	1	1	2	0	1	1	2	0	**
2159	2159007	1	1	0	2	1	1	2	0	0	2	**
2159	2159008	2	0	0	2	2	0	1	1	2	0	**
2159	2159009	2	0	0	2	2	0	2	0	2	0	**
2159	2159010	2	0	0	2	2	0	1	1	2	0	**
2159	2159012	2	0	0	2	1	1	2	0	2	0	**
2159	2159013	2	0	0	2	2	0	1	1	2	0	**
2159	2159016	1	1	1	1	1	1	1	1	1	1	**
2159	2159017	2	0	0	2	2	0	2	0	2	0	**
2163	2163001	1	1	0	2	0	2	0	2	0	2	**
2163	2163002	0	2	0	2	0	2	0	2	0	2	**
2163	2163003	1	1	0	2	0	2	0	2	0	2	**
2163	2163005	1	1	0	2	0	2	0	2	0	2	**
2163	2163006	0	2	0	2	1	1	0	2	0	2	**
2163	2163007	0	2	0	2	0	2	0	2	0	2	**
2163	2163009	2	0	0	2	0	2	0	2	0	2	**
2164	2164002	0	2	0	2	0	2	0	2	0	2	**
2164	2164003	0	2	0	2	0	2	0	2	0	2	**
2164	2164004	0	2	0	2	0	2	0	2	0	2	**
2164	2164005	0	2	0	2	0	2	0	2	0	2	**
2165	2165001	0	2	0	2	0	2	0	2	0	2	**
2165	2165002	0	2	0	2	0	2	0	2	0	2	**
2165	2165003	1	1	0	2	0	2	0	2	0	2	**
2165	2165004	0	2	0	2	0	2	0	2	0	2	**
2165	2165005	1	1	0	2	0	2	0	2	0	2	**
2165	2165006	1	1	0	2	0	2	0	2	0	2	**
2165	2165007	1	1	0	2	0	2	0	2	0	2	**
2165	2165008	1	1	0	2	0	2	0	2	0	2	**
2165	2165011	0	2	0	2	0	2	0	2	0	2	**
2165	2165012	0	2	0	2	1	1	0	2	0	2	**
2165	2165013	0	2	0	2	0	2	0	2	*	*	**
2165	2165014	0	2	0	2	0	2	0	2	0	2	**
2165	2165015	0	2	0	2	0	2	1	1	0	2	**
2165	2165016	0	2	0	2	0	2	0	2	0	2	**
2165	2165017	0	2	0	2	0	2	0	2	0	2	**
2165	2165018	0	2	0	2	0	2	0	2	0	2	**
2165	2165019	1	1	0	2	1	1	0	2	0	2	**
2165	2165020	0	2	0	2	0	2	1	1	0	2	**
2166	2166001	2	0	0	2	2	0	2	0	2	0	**
2166	2166003	2	0	0	2	2	0	2	0	0	2	**
2166	2166005	2	0	0	2	2	0	2	0	2	0	**
2166	2166006	1	1	0	2	1	1	1	1	*	*	**
2166	2166007	2	0	1	1	2	0	2	0	2	0	**
2166	2166008	2	0	0	2	2	0	2	0	2	0	**
2166	2166009	2	0	1	1	2	0	2	0	2	0	**
2166	2166011	2	0	0	2	2	0	2	0	2	0	**
2166	2166012	2	0	1	1	2	0	1	1	2	0	**
2166	2166014	2	0	0	2	2	0	2	0	2	0	**
2166	2166015	2	0	0	2	2	0	2	0	2	0	**
2166	2166016	2	0	0	2	2	0	2	0	0	2	**
2166	2166017	2	0	1	1	2	0	1	1	2	0	**
2166	2166019	1	1	0	2	2	0	2	0	2	0	**
2166	2166020	2	0	0	2	2	0	2	0	2	0	**
2166	2166021	2	0	2	0	*	*	2	0	2	0	**
2166	2166022	2	0	2	0	2	0	1	1	2	0	**
2166	2166023	2	0	0	2	2	0	2	0	2	0	**
2166	2166024	2	0	0	2	2	0	2	0	2	0	**
2166	2166025	1	1	1	1	2	0	2	0	2	0	**
2166	2166026	2	0	1	1	2	0	1	1	2	0	**
2166	2166027	2	0	0	2	2	0	2	0	2	0	**
2166	2166028	2	0	0	2	2	0	2	0	2	0	**
2166	2166029	2	0	0	2	2	0	2	0	2	0	**
2166	2166030	2	0	0	2	2	0	2	0	2	0	**
2166	2166031	1	1	1	1	2	0	2	0	2	0	**
2166	2166032	2	0	0	2	2	0	1	1	2	0	**
2166	2166034	2	0	1	1	2	0	2	0	2	0	**
2166	2166035	1	1	2	0	2	0	2	0	2	0	**
2166	2166036	2	0	2	0	2	0	2	0	2	0	**
2166	2166037	2	0	0	2	2	0	2	0	2	0	**
2166	2166038	1	1	0	2	2	0	2	0	2	0	**
2166	2166039	1	1	0	2	2	0	2	0	2	0	**
2166	2166041	2	0	1	1	2	0	2	0	2	0	**
2166	2166042	2	0	0	2	2	0	0	2	2	0	**
2166	2166043	2	0	0	2	2	0	2	0	2	0	**
2166	2166044	2	0	0	2	2	0	2	0	2	0	**
2166	2166045	2	0	1	1	2	0	2	0	0	2	**
2166	2166047	2	0	0	2	2	0	2	0	2	0	**
2166	2166049	2	0	0	2	2	0	1	1	2	0	**
2166	2166050	2	0	0	2	2	0	2	0	2	0	**
2166	2166051	2	0	2	0	2	0	2	0	2	0	**
2167	2167001	1	1	1	1	2	0	2	0	2	0	**
2167	2167002	2	0	0	2	2	0	2	0	2	0	**
2167	2167003	2	0	1	1	2	0	2	0	2	0	**
2167	2167004	2	0	0	2	2	0	2	0	2	0	**
2200	2200001	0	2	0	2	0	2	0	2	0	2	**
2200	2200002	2	0	0	2	0	2	0	2	0	2	**
2200	2200003	0	2	0	2	0	2	0	2	0	2	**
2200	2200010	0	2	0	2	1	1	0	2	0	2	**
2200	2200011	0	2	0	2	0	2	0	2	0	2	**
2200	2200012	0	2	0	2	0	2	0	2	0	2	**
2200	2200013	0	2	0	2	0	2	0	2	0	2	**
2200	2200014	0	2	0	2	0	2	0	2	*	*	**
2200	2200015	0	2	0	2	0	2	0	2	2	0	**
2200	2200016	0	2	0	2	0	2	0	2	0	2	**

Appendices for Chapter 3

Appendix 3.1. overleaf. Ecological variables measured at each site. The site name is given on the first page only. The first column of numbers corresponds to the site on subsequent pages. For a definition of the variables see Chapter 3.

	Site	mean p(v)	Aquatic Habitat	Habitat	D.Function	Width(m)	length	depth	Max Bank depth	Bank incline
1	1001	0.730	wheelrut	2	-1.08	1.20	4.50	0.22	0.13	shallow
2	1002	0.684	wheelrut	2	0.15	0.88	14.60	0.16	0.10	
3	1003	0.740	wheelrut	2	-0.82	2.00	12.00	0.20	0.10	
4	1004	0.455	depression	1	1.41	2.40	3.80	0.21	0.06	
5	1005	0.125	drainage canal	1	2.06	1.90	100.00	0.51	0.10	
6	1010	0.250	drainage canal	0		1.50	100.00			
7	1011	0.381	wheelrut	0		0.36	100.00	0.16	0.12	
8	1013	0.181		0						
9	1014	0.023	depression	1	1.95	2.70	0.21	0.50	0.00	
10	1016	0.042	wheelrut	2	-0.52	3.70	5.00	0.17	0.17	
11	1018	0.500		0						
12	1019	0.227	wheelrut	2	-0.71	1.00	1.40	0.12	0.10	steep
13	1025	1.000		0						
14	1028	0.983		0						
15	1029	0.984	wheelrut	2	-1.60	0.40	4.00	0.06	0.25	
16	1032	0.000	wheelrut	0		0.30	3.00			
17	1033	0.069	wheelrut	2	-1.01	0.30	6.00	0.12	0.18	steep
18	1035	0.080	depression	1	3.03	10.00	15.00	0.30	0.00	shallow
19	1036	0.062	depression	1	1.84	13.50	15.50	0.30	0.00	shallow
20	1037	0.050	pig trample	0				0.05		
21	1038	0.367	drainage canal	2	-1.28	0.50	100.00	0.09	1.50	steep
22	1039	0.062	pond	1	4.82	25.00	38.00	0.70	0.10	none
23	1040	0.068	pond	1	1.96	6.00	35.00	0.35	5.00	none
24	1041	0.600	drainage canal	0						
25	1042	0.056	drainage canal	1	2.42	2.40	100.00	0.27	0.20	medium
26	1043	0.126	depression	1	3.93	12.00	19.50	0.13	0.00	none
27	1044	0.236	wheelrut	2	0.89	3.20	20.00	0.09	0.80	medium
28	1045	0.087	drainage canal	1	2.81	2.60	100.00	0.16	0.00	none
29	1046	0.458	wheelrut	2	-0.90	0.40	10.00	0.20	0.25	steep
30	1047	0.437	wheelrut	2	0.82	0.90	2.40	0.80	0.15	medium
31	1049	0.640	wheelrut	2	0.04	0.60	100.00	0.10	0.20	steep
32	1050	0.039	drainage canal	1	2.68	2.00	100.00	0.55	0.10	medium
33	1051	0.150	wheelrut	2	0.05	1.50	35.00	0.25	0.15	none to steep
34	1052	0.063	pond	1	3.01	25.00	30.00	1.00	0.30	none to steep
35	1053	0.037	pond	1	4.75	15.00	30.00	1.00	0.05	steep
36	1054	0.594	wheelrut	2	-1.08	0.50	21.60	0.20	0.10	steep
37	1055	0.139	pond	1	3.95	10.00	15.00	0.70	0.10	shallow
38	1056	0.218	furrow	2	-0.54	2.00	52.00	0.22	0.25	none to steep
39	1057	0.000	wheelrut	2	-0.82	0.80	17.00	0.15	0.20	steep
40	1058	1.000	wheelrut	0						
41	1059	0.767	wheelrut	0						
42	1060	0.750	furrow	0						
43	1061	0.187	wheelrut	0						
44	1063	0.301	village canal	1	3.98	15.00	80.00	1.00	0.00	none
45	1064	0.252	furrow	2	0.18	0.50	100.00	0.18	0.03	steep
46	1066	0.219	depression	2	0.27	3.00	3.00	0.17	10.00	medium
47	1067	0.583	drainage canal	0						
48	1068	0.625	drainage canal	0						
49	1069	0.150	village canal	2	0.57	0.90	0.30	0.28	0.20	
50	1070	0.819	wheelrut	2	-0.58	0.80	9.00	0.10	0.07	
51	1071	0.900	wheelrut	2	-1.23	0.35	7.00	0.12	0.10	
52	1072	0.875	wheelrut	2	-1.08	2.00	13.50	0.18	0.12	
53	1073	0.875	wheelrut	0						
54	1074	0.850	wheelrut	2	-1.30	0.60	100.00	0.18	0.20	
55	1075	0.875	village canal	0						
56	1076	0.938	wheelrut	0						
57	1077	0.889	wheelrut	2	-1.11	1.50	2.00	0.09	0.06	
58	1078	0.781	wheelrut	2	0.53	2.00	9.00	0.28	0.13	
59	1079	0.833	wheelrut	0						
60	1080	0.750	wheelrut	2	0.23	0.60	2.50	0.20		
61	1081	0.125	drainage canal	1	1.93	2.00	100.00	0.54	0.10	
62	1082	0.250	wheelrut	0		0.60		0.10		
63	1083	0.625	furrow	0		1.00	5.00			

	Site	mean p(v)	Aquatic Habitat	Habitat	D.Function	Width(m)	length	depth	Max Bank depth	Bank incline
64	1084	0.250	furrow	2	-0.51	3.50	6.00	0.35	0.00	none
65	1085	0.417	wheelrut	0						
66	1086	0.750	wheelrut	0						
67	1087	0.825	wheelrut	0						
68	1089	1.000	wheelrut	2	-0.94	2.50	3.00	0.25	0.20	
69	1091	0.938	wheelrut	2	-0.92	2.00	3.00	0.12	0.11	
70	1092	0.875	wheelrut	0						
71	1097	0.854	wheelrut	2	-1.63	0.40	10.00	0.05	0.20	
72	1098	1.000	wheelrut	0						
73	1099	0.792	wheelrut	2	-1.18	0.60	2.50	0.08	0.15	
74	1100	0.833	wheelrut	2	-0.33	1.00	5.00	0.18	0.15	
75	1103	0.227	wheelrut	2	-0.95	0.70	7.00	0.10	0.17	
76	1104	0.202	depression	1	2.05	8.00	10.00	0.08	0.00	
77	1105	0.167	wheelrut	2	-1.24	0.90	100.00	0.16	0.15	
78	1109	0.000	furrow	2	-1.59	0.60	35.00	0.06	0.00	none
79	1110	0.577	wheelrut	2	-0.46	0.80	100.00	0.15	0.20	
80	1111	1.000	wheelrut	2	-1.24	0.60	3.00	0.13	0.10	
81	1112	0.854	wheelrut	2	-1.51	0.60	5.00	0.08	0.20	
82	1113	0.609	wheelrut	2	-1.31	0.85	100.00	0.15	0.13	
83	2012	0.464	wheelrut	2	-0.46	0.80	100.00	0.15	0.20	
84	2054	0.125		0						
85	2115	0.093	drainage canal	2	0.06	1.50	100.00	0.35	0.20	
86	2116	0.081	pond	1	3.04	12.00	35.00	0.30	0.00	
87	2117	0.074	drainage canal	2	-1.10	1.16	6.00	0.25	0.70	steep
88	2118	0.813	wheelrut	0						
89	2119	0.071	wheelrut	2	-1.40	0.40	15.00	0.12	0.20	
90	2120	0.114	wheelrut	2	-0.23	0.35	20.00	0.14	0.30	
91	2121	0.020	drainage canal	2	0.08	2.20	6.00	0.20	0.50	steep
92	2122	0.938	wheelrut	2	-1.40	0.30	2.60	0.11	0.15	
93	2124	0.813	wheelrut	2	-1.46	0.40	1.25	0.10	0.01	
94	2126	0.828	wheelrut	2	-1.29	0.40	4.00	0.15	0.20	
95	2127	0.969	wheelrut	2	-1.41	0.30	3.50	0.14	0.15	
96	2132	0.938	wheelrut	2	-1.45	0.45	1.80	0.12	0.06	
97	2133	0.922	wheelrut	2	-1.55	0.30	1.20	0.06	0.03	
98	2134	0.886	wheelrut	2	-1.45	0.30	1.70	0.05	0.06	
99	2135	0.095	wheelrut	2	-1.07	0.60	2.30	0.18	0.05	steep
100	2136	0.900	wheelrut	2	-1.55	0.40	5.00	0.08	0.20	steep
101	2138	0.875	wheelrut	2	-1.41	0.35	4.00	0.14	0.15	steep
102	2140	0.773	wheelrut	0						
103	2141	0.846	wheelrut	0						
104	2142	0.167	pig trample	2	-1.67	0.35	2.00	0.02	0.43	steep
105	2143	0.134	wheelrut	2	-1.12	0.58	4.00	0.18	0.20	steep
106	2144	1.000	wheelrut	0						
107	2145	0.208	wheelrut	2	-1.50	0.45	2.50	0.10	0.20	steep
108	2146	0.792	wheelrut	2	-1.37	0.35	100.00	0.12	0.50	steep
109	2147	0.449	wheelrut	2	-1.41	0.40	20.00	0.14	0.10	steep
110	2148	0.125	wheelrut	0						
111	2149	0.500		0						
112	2150	0.222		0						
113	2151	0.097	drainage canal	2	-1.04	0.60	13.00	0.25		
114	2152	0.094	pond	1	4.44	30.00	100.50	0.30	0.00	
115	2153	0.250	depression	1	6.70	40.00	100.00	0.40	0.00	
116	2154	0.755	wheelrut	2	-1.38	0.30	2.00	0.08	0.20	steep
117	2155	0.813	well	2	-1.46	0.40	0.80	0.10	0.25	steep
118	2156	0.854	wheelrut	2	-1.40	0.60	6.50	0.14	0.30	
119	2157	0.125	wheelrut	1	1.28	0.30	2.50	0.15	0.10	medium steep
120	2158	0.850	wheelrut	2	-1.55	0.40	5.00	0.08	0.00	
121	2159	0.134	pond	1	5.29	15.00	30.00	0.50	0.00	
122	2163	0.893	wheelrut	2	-1.23	0.50	2.50	0.12	0.15	shallow-steep
123	2164	1.000	well	0		3.00	3.00		0.50	
124	2165	0.930	wheelrut	2	-0.72	2.00	8.00	0.24	0.50	steep
125	2166	0.069	drainage canal	2	0.42	0.90	20.00	0.12	0.50	steep
126	2167	0.031	drainage canal	1	4.70	3.00	30.00	0.15	1.00	shallow-steep

	% Tree Cover	% Ev	• %Sv	Sh Veg0-15	Sh Veg15-50	Sh veg>50	pH soil	pH water	Temp air	Temp water	Altitude
1	0	0	1	5	5	0	6.5		11.0	11.0	103
2	20	1	0	80	5	0	6.0	5.5	13.0	14.0	104
3	0	0	0	20	10	0	6.5	7.5	16.0	17.0	104
4	0	20	0	90	0	0	6.5		10.0	11.0	101
5	40	50	0	10	80	10	6.5		10.0	11.0	99
6									13.0	13.0	99
7	40	0	0				6.0	6.5	14.0	14.0	99
8				20	20	30	6.5	7.0	11.0	12.0	99
9	0	20	10	80	20	0	7.0	8.0	14.0	16.0	99
10	0	10	5	10	10	0	6.5	6.0	10.0	11.5	99
11											99
12	0	0	0	40	0	0	7.0	6.5	9.5	10.5	99
13											200
14											200
15	0	0	0	0	0	0	6.0	6.8	24.5	24.0	200
16											97
17	10	2	0	20	10	0	5.0	6.0	18.0	15.5	97
18	0	40	100	95	5	0	6.5	7.5	15.0	21.0	100
19	0	20	80	70	25	5	6.5	7.7	15.5	19.0	100
20		0	0								99
21	20	5	1	1	30	10	5.0	6.0	18.0	20.0	99
22	5	70	10	20	20	60	6.5	7.7	19.0	20.0	98
23	15	40	10	40	0	60	6.7	7.0	18.5	21.0	100
24											
25	5	70	100	10	30	60	7.0	8.5	18.5	18.0	100
26	0	70	100	85	15	0	6.5	7.3	17.0	18.0	100
27	0	20	100	75	20	5	6.0	5.7	15.5	20.0	100
28	2	80	100	20	70	10	6.0	5.5	11.5	15.0	100
29	10	0	0	20	40	10	5.5	5.5	12.0	15.0	100
30	0	5	0	20	20	0	5.7	6.5	18.0	21.5	100
31	20	2	1	80	20	0	6.0	6.5	15.0	18.0	101
32	0	30	90	90	0	0	7.5	7.3	23.9	18.4	100
33	0	5	0	50	0	0	5.5	6.5	18.0	21.0	100
34	25	15	20	20	0	0	6.0	6.5	18.0	19.0	100
35	25	60	10	30	40	20	6.5	7.0	13.0	16.5	
36	10	0	0	10	0	0	6.5	6.3	17.0	18.0	100
37	10	70	100	15	15	70		6.7	12.0	17.5	100
38	0	5	0	20	30	10	6.0	6.0	16.0	17.0	100
39	0	0	0	30	30	0	6.5	6.7	16.0	16.0	99
40											101
41											103
42											100
43											100
44	0	20	90	90	10	0	6.7	7.7	19.0	22.0	100
45	0	5	2	70	30	0	5.5	7.0	17.5	21.0	100
46	0	30	60	5	85	10	6.0	6.7	28.0	23.0	99
47											
48											
49	2	35	10	0	50	50	6.6	7.2	23.5	18.0	100
50	0	0	1	50	0	0	6.0	6.8	19.0	26.0	115
51	0	1	0	10	0	0	5.0	5.0	20.0	26.5	115
52	80	2	0	3	10	0	6.0	6.0	19.0	20.0	112
53											115
54	0	0	0	0	0	0	5.5	6.0	25.5	25.0	126
55											108
56											112
57	0	0	0	20	0	0	6.8	7.0	20.0	25.5	112
58	0	25	0	20	75	5	5.0	5.0	24.0	20.0	115
59											112
60	70	1	0	80	20	0					112
61	0	30	80	50	20	30	7.0	7.2	28.0	22.0	100
62		0	0								100
63											104

	% Tree Cover	% Ev	%SV	Sh Veg0-15	Sh Veg15-50	Sh veg>50	pH soil	pH water	Temp air	Temp water	Altitude
64	0	5	5	0	100	0	6.0	6.5	22.5	23.0	104
65		0	0								104
66											104
67											115
68	100	0	5	5	0	0	6.0	6.5	12.5	10.5	112
69	0	0	0	25	3	0	5.5	5.5	12.5	12.0	112
70		20									124
71	50	0	0	0	0	0	5.5	5.5	14.0	12.5	124
72											114
73	20	0	0	20	0	0	6.0	6.0	13.0	14.5	114
74	20	5	0	40	0	0	6.3	7.0	15.0	17.5	114
75	25	0	0	30	2	0	6.0	6.5	15.0	14.0	99
76	30	50	5	50	3	2	6.0	7.0	18.0	19.0	99
77	10	0	0	5	3	5	6.0	6.0	14.0	15.0	99
78	0	0	2	0	0	0	6.0	7.0	22.5	26.0	100
79	20	0	0	50	0	30	6.3	6.5	19.0	19.0	99
80	10	0	0	10	0	0	5.7	5.7	22.5	19.5	104
81	30	0	0	2	0	0	5.5	6.5	22.0	25.0	104
82	0	0	0	3	0	0	6.0	6.8	25.0	24.0	99
83	20	0	0	50	0	30	6.3	6.5	19.0	19.0	99
84											
85	35	0	50	50	20	0		6.0	22.0	20.0	100
86	0	75	10	0	50	50		6.5	20.0	21.0	100
87	55	0	0	0	0	0		6.0	16.0	15.0	99
88				2	0	0		7.6	18.0	15.0	112
89	0	0	0	3	7	0		7.8	19.0	19.0	97
90	0	10	5	40	5	0		7.9	19.5	21.0	97
91	0	25	10	5	70	10		6.0	15.5	17.0	97
92	95	0	0	5	30	0		7.5	15.0	13.5	209
93	90	1	1	0	0	0		7.5	15.5	14.0	209
94	50	0	0	5	5	0		7.6	16.0	15.0	209
95	10	0	0	0	0	0		7.3	17.0	17.0	209
96	50	0	0	0	0	0		6.0	18.5	23.0	209
97	5	0	0	3	10	0		5.0	18.0	19.0	209
98	30	0	0	10	0	0		5.0	17.0	18.0	209
99	0	5	5	0	0	0		6.0		21.0	97
100	30	0	0	0	40	5		5.0	17.5	15.0	209
101	95	0	0	0	0	0		6.0	18.0	18.0	209
102											209
103											209
104	50	0	0	2	20	0		7.0	22.0	21.0	98
105	0	0	3	10	70	0		6.0	22.5	20.0	98
106											98
107	40	0	65	0	90	0		6.0	24.0	22.0	98
108	60	0	0	5	0	0		6.5	19.0	18.0	100
109	80	0	0	0	0	0		6.5	18.0	18.5	100
110											98
111											97
112											
113	0	2	0	0	60	40		6.5	24.5	21.0	100
114	10	85	80	0	0	100		6.0	21.5	21.5	98
115	0	98	100	50	50	0		6.0	24.0	24.0	102
116	95	0	0	10	2	0		6.0	20.0	18.0	127
117	0	1	0	0	100	0		6.0		18.5	127
118	95	0	0	0	0	0		6.0	20.0	19.5	127
119	10	50	30	20	80	0		6.0	24.0	23.0	114
120	60	0	0	0	0	0		7.0	24.0	24.0	114
121	3	100	50	25	25	50					114
122	85	1	0	10	0	0		6.0	19.0	17.0	187
123	0	5	0	100	0	0		6.0	20.0	19.0	280
124	50	0	0	20	0	0		6.5	19.0	18.0	280
125	15	3	2	95	0	0		7.0	24.0	22.0	100
126	80	90	10	100	0	0		7.0	24.0	24.0	100

	Turbidity	Pond substrate	Soil type	Immediate area	Surround	region	Position
1	clear	mud	clay	forest floor	forest/arable edge	Pescenica	lowland
2	clear	leaf litter	clay	forest floor	forest	Pescenica	lowland
3	clear	mud	clay	clearway	forest	Pescenica	lowland
4	clear	leaf litter	clay	clearing	forest	Bulge	lowland
5	21cm	rotting veg	clay	wetland	forest	Bulge	lowland
6				forest floor	forest	Bulge	lowland
7	murky	mud	clay	clearway	forest	Bulge	lowland
8	clear			clearway	forest/arable edge		
9	clear	mud	clay	wetland	forest/arable edge	Bulge	lowland
10	clear	leaf litter	dark loose soil	clearway	forest	Bulge	lowland
11				clearway	forest	Bulge	lowland
12	clear	mud	clay	clearway	forest/arable edge	Bulge	lowland
13						Kupa	highland
14						Kupa	highland
15	1cm	mud	clay	clearway	forest	Kupa	highland
16		leaf litter		clearway	forest/arable edge	Lekenik	lowland
17	almost clear	leaf litter	light clay	clearway	forest/arable edge	Lekenik	lowland
18	clear	veg	clay	wetland	arable	Lekenik	lowland
19	clear	mud	clay	wetland	arable	Lekenik	lowland
20		mud	clay	clearway	forest	Bulge	lowland
21		rotting veg	compact clay	forest floor	forest	Bulge	lowland
22	20cm	rotting veg	clay	wetland	arable	Velesevec etc	lowland
23	clear	rotting veg	clay	wetland	arable	Velesevec etc	lowland
24				clearway	forest	Bulge	lowland
25	clear	veg	clay	field	arable	Lekenik	lowland
26	clear	veg	clay	wetland	arable	Lekenik	lowland
27	clear	veg	heavy clay	wetland	forest/arable edge	Lekenik	lowland
28	clear	veg	clay	wetland	forest/arable edge	Lekenik	lowland
29	15cm	mud	clay	forest floor	forest/arable edge	Lekenik	lowland
30	clear	mud	clay	wetland	forest/arable edge	Lekenik	lowland
31	clear	rotting veg	clay	clearway	forest	Bulge	lowland
32	clear	rotting veg	light clay	pasture	arable	Velesevec etc	lowland
33	15cm	mud	light clay	pasture	arable	Velesevec etc	lowland
34	50cm	mud	clay	wetland	arable	Velesevec etc	lowland
35	clear	veg	light clay	wetland	arable	Velesevec etc	lowland
36	10cm	mud	clay	field	forest/arable edge	Lekenik	lowland
37	clear	veg	clay	field	arable	Lekenik	lowland
38	7cm	mud	clay	field	forest/arable edge	Lekenik	lowland
39	6cm	mud	heavy clay	clearway	forest/arable edge	Bulge	lowland
40				clearway	forest	Bulge	lowland
41				clearing	forest/arable edge	Pescenica	lowland
42				field	arable	Lekenik	lowland
43				field	arable	Lekenik	lowland
44	10cm	veg	clay	pasture	residential	Lekenik	lowland
45	12cm	mud	clay	field	arable	Lekenik	lowland
46	clear	rotting veg	clay	wetland	forest/arable edge	Beyond Domingo	lowland
47				field	forest/arable edge	Pescenica	lowland
48				field	arable	Pescenica	lowland
49	clear	rotting veg	clay	tarmac	residential	Lekenik	lowland
50	7cm	mud	clay	field	forest/arable edge	Polijana	midland
51	2cm	mud	clay	clearing	forest/arable edge	Polijana	midland
52	clear	mud	clay	forest floor	forest/arable edge	Pescenica	lowland
53		mud	clay	clearing	forest/arable edge	Polijana	midland
54	8cm	mud	clay	clearing	forest/arable edge	Polijana	midland
55				tarmac	residential	Pescenica	lowland
56				field	forest/arable edge	Pescenica	lowland
57	1cm	mud	clay	field	forest/arable edge	Pescenica	lowland
58	5cm	rotting veg	clay	forest floor	forest/arable edge	Polijana	midland
59				clearing	forest/arable edge	Pescenica	lowland
60	clear	leaf litter		clearing	forest/arable edge	Pescenica	lowland
61	clear	rotting veg	clay	wetland	forest/arable edge	Beyond Domingo	lowland
62				wetland	forest/arable edge	Beyond Domingo	lowland
63				forest floor	forest/arable edge	Polijana	midland

	Turbidity	Pond substrate	Soil type	Immediate area	Surround	region	Position
64	clear	mud	clay	field	arable	Polijana	midland
65		mud		field	arable	Polijana	midland
66				clearway	forest/arable edge	Polijana	midland
67				clearing	forest	Polijana	midland
68	clear	rotting veg	clay	forest floor	forest/arable edge	Pescenica	midland
69	7cm	mud	clay	clearing	forest	Pescenica	midland
70				clearing	forest	Brodic	midland
71	1cm	mud	clay	forest floor	forest	Brodic	highland
72				clearway	forest	Pescenica	lowland
73	2cm	mud	clay	forest floor	forest/arable edge	Pescenica	lowland
74	clear	mud	clay	forest floor	forest/arable edge	Pescenica	lowland
75	clear	mud	clay	clearway	forest	Bulge	lowland
76	clear	rotting veg	clay	clearway	forest/arable edge	Bulge	lowland
77	clear	mud	clay	clearway	forest/arable edge	Bulge	lowland
78	clear	rotting veg	clay	field	arable	Lekenik	lowland
79	10cm	mud	clay	clearway	forest	Bulge	lowland
80	7cm	mud	clay	clearway	forest	Bulge	lowland
81	1cm	mud	clay	clearing	forest	Bulge	lowland
82	clear	mud	clay	clearway	forest	Bulge	lowland
83	10cm	mud	clay	clearway	forest		
84						Lekenik	lowland
85	clear	mud		wetland	forest/arable edge	Lekenik	lowland
86	slightly murky	rotting veg	clay	wetland	arable	Velesevec etc	lowland
87	turbid	mud	clay	clearway	forest/arable edge	Bulge	lowland
88	clear			forest floor	forest	Polijana	midland
89	slightly murky	mud		clearing	forest	Beyond Domingo	
90	clear	mud	crumbly clay	clearing	forest	Beyond Domingo	lowland
91	murky	rotting veg	crumbly clay	clearing	forest	Beyond Domingo	lowland
92	clear	leaf litter		forest floor	forest	Perkovec	
93	murky	mud		forest floor	forest	Perkovec	
94	murky	mud		forest floor	forest	Perkovec	
95	murky	mud		forest floor	forest	Perkovec	
96	murky	mud		forest floor	forest	Perkovec	
97	clear	leaf litter		forest floor	forest	Perkovec	lowland
98	slightly murky	mud		forest floor	forest	Perkovec	
99	slightly murky	mud		forest floor	forest	Beyond Domingo	lowland
100	slightly murky	mud		forest floor	forest	Perkovec	highland
101	murky	mud		forest floor	forest	Perkovec	highland
102				forest floor	forest	Perkovec	highland
103				forest floor	forest	Perkovec	highland
104	slightly murky	mud		clearway	forest/arable edge	Memorial forest	highland
105	slightly murky	mud		clearway	forest/arable edge	Memorial forest	lowland
106				forest floor	forest	Memorial forest	highland
107	slightly murky	leaf litter		forest floor	forest	Memorial forest	highland
108	murky	mud		clearway	forest	Memorial forest	highland
109	murky	mud		forest floor	forest	Memorial forest	highland
110				clearway	forest	Memorial forest	lowland
111							highland
112							highland
113	clear	mud		tarmac	residential	Lekenik	highland
114	clear	mud		forest floor	forest	Memorial forest	
115	clear	veg		pasture	arable	Zazina	lowland
116	slightly murky	mud		forest floor	forest	Duzica	
117	slightly murky	mud		clearway	forest	Duzica	lowland
118	slightly murky	mud	clay	forest floor	forest	Duzica	lowland
119	clear	rotting veg		wetland	forest/arable edge	Duzica	
120	slightly murky	mud		forest floor	forest/arable edge	Duzica	lowland
121	clear	veg		field	arable	Duzica	lowland
122	slightly murky	mud		forest floor	forest	Brodic	
123	clear	?		pasture	forest/vineyard	Brodic	
124	murky	mud		forest floor	forest/vineyard	Brodic	lowland
125	slightly murky	mud		pasture	arable	Greda	lowland
126	clear	veg		pasture	arable	Greda	lowland

Appendix 3.2. Notes from Szymura describing the 20 sites he sampled in 1979. On the basis of these notes each site was assigned to a habitat type; 1=pond, 2=puddle, 0 was assigned where it was undecided. Note the different site numbering between Szymura's and the present study. Those sites with a defined habitat were used to describe the cline in Chapter 3, section 4.

Szymura's site no	Present Site no	Annotated description of site from Szymura's notes (nearest place name in brackets)	Assigned to habitat
1	1	wide shallow ponds, also in small ditches (Jezevo)	1
2	2	old ditches dug out in peaty meadow, lots of vegetation (Cret)	1
3	3	Large pond; vegetation rich (Orlé)	1
4	4	Pond and ditches, vegetation rich, (Stružec).	1
5	5	Pond in a depression in pasture (Selce/Odra)	1
6	6	Old ditches with vegetation (Vratova)	1
7	7	Ditches along road, mixed habitat (Stanci)	0
8	8	Wheelruts on road, some vegetaion (Pešćenica forest).	2
9a	9	Puddles on a narrow forest road, no vegetation (Prekobunje)	2
9b	10	Ditch with rushes and grasses surrounded by forest (Prekobunje, Buna).	1
10a	11	Small ditch, forest edge, clear water,some vegetation (Turopolje).	2
10b	12	Ditches and tractor wheelruts on road (Turopolje forest).	2
11	13	Small ponds in wheelruts, forest edge (Buševac)	2
12	14	Small ponds, village road, little vegetation, typical variegata site (Pešćenica).	2
13	15	Slowly flowing forest stream, nearby spring, wheelruts, some vegetation (Vukojevac).	0
14	16	Drying out stream on clay substrate, some vegetation (Vukojevac).	0
15	17	Muddy puddles, forest edge, no vegetaion (Modruše).	2
16	18	Puddles on clay, no vegetation (Severje)	2
17	19	Wheelruts on forest road (Cerje)	2
18	20	Pond close to vineyards on hills, grasses at edge (Perkovec).	2

Appendix 3.3.A

The most likely cline with variable width and α fitted in two dimensions. The cline is described by nine segments each 4km in length. The total likelihood of this cline is LogL=-145.58. The total number of sites used to describe the cline is 134. These are sites with known habitat type.

Table legend:- **p(obs)** is the observed gene frequency of the site ($=\bar{p}$)
Pexp is the expected frequency given the model.
hab is the habitat type:1=ponds, 2=puddles, 0=unknown
LnL; the likelihood of the model for a particular site; if >2 p<0.05
Dist is the distance of that site from the centre of the cline (km)
S is the distance along the cline from the starting point
Width is the width of the cline at distance S along the cline
Seg is the nearest segment to the site.
Dist/W is distance of a site from the centre of the cline divided by the width at that point. This co-ordinate was used to reduce the cline to one dimension.

Site	p(obs)	hab	Pexp	lnL	Dist	S	Width	Seg	Dist/W
1	0.079	1	0.049	-1.487	-18.22	4.00	6.73	2	-2.71
2	0.067	1	0.068	-0.003	-8.50	4.00	6.73	2	-1.26
3	0.059	1	0.061	-0.003	-11.72	4.00	6.73	2	-1.74
4	0.068	1	0.067	-0.002	-5.86	18.00	4.34	5	-1.35
5	0.121	1	0.076	-1.240	-3.03	17.00	3.87	5	-0.78
6	0.121	1	0.261	-6.756	-0.77	16.92	3.84	5	-0.20
7	0.528	0	0.464	0.000	0.68	16.56	3.67	5	0.18
8	0.800	2	0.723	-0.828	1.40	17.28	4.00	5	0.35
9	0.800	2	0.769	-0.078	1.91	16.00	3.41	5	0.56
10	0.286	1	0.077	-3.809	-2.35	16.00	3.41	5	-0.69
11	0.042	1	0.075	-0.232	-1.29	9.35	1.53	3	-0.84
12	0.650	2	0.818	-1.516	5.66	4.00	6.73	2	0.84
13	0.604	0	0.449	0.000	0.22	9.13	1.57	3	0.14
14	0.585	2	0.770	-3.113	2.30	19.58	4.07	5	0.57
15	0.932	2	0.866	-6.170	8.08	4.00	6.73	2	1.20
16	0.781	0	0.742	0.000	4.87	20.00	3.99	6	1.22
17	0.918	2	0.888	-0.906	5.66	20.00	3.99	6	1.42
18	0.885	2	0.861	-0.352	4.61	20.00	3.99	6	1.15
19	0.919	2	0.916	-0.009	7.04	20.00	3.99	6	1.76
20	0.955	2	0.904	-1.910	11.51	2.00	7.22	1	1.59
1001	0.730	2	0.786	-1.135	1.28	12.86	1.97	4	0.65
1002	0.684	2	0.669	-0.048	0.26	12.21	1.75	4	0.15
1003	0.740	2	0.697	-0.565	0.45	12.44	1.83	4	0.25
1004	0.455	1	0.572	-0.222	0.68	16.56	3.67	5	0.18
1005	0.125	1	0.318	-0.507	-0.50	16.87	3.82	5	-0.13
1010	0.250	0	0.246	0.000	-0.30	16.00	3.41	5	-0.09
1011	0.381	2	0.391	-0.011	-0.57	16.00	3.41	5	-0.17
1013	0.181	0	0.057	0.000	-1.60	16.25	3.52	5	-0.45
1014	0.023	1	0.079	-3.017	-2.41	17.30	4.01	5	-0.60
1016	0.042	2	0.372	-7.831	-0.64	16.00	3.41	5	-0.19
1018	0.500	0	0.071	0.000	-1.37	16.00	3.41	5	-0.40
1019	0.227	2	0.149	-0.321	-1.83	16.72	3.74	5	-0.49

Appendix 3.3A continued.

Site	p(obs)	hab	Pexp	lnL	Dist	S	Width	Seg	Dist/W
1025	1.000	2	0.969	-0.287	21.39	2.00	7.22	1	2.96
1028	0.983	2	0.969	-0.283	21.39	2.00	7.22	1	2.96
1029	0.984	2	0.969	-1.364	21.39	2.00	7.22	1	2.96
1032	0.000	0	0.031	0.000	-6.20	18.00	4.34	5	-1.43
1033	0.069	2	0.111	-0.814	-3.94	19.49	4.08	5	-0.97
1035	0.080	1	0.078	-0.003	-2.40	20.32	3.76	6	-0.64
1036	0.062	1	0.078	-0.025	-2.55	20.14	3.89	6	-0.66
1037	0.050	0	0.161	0.000	-0.60	15.16	2.97	4	-0.20
1038	0.367	2	0.401	-0.022	-0.45	14.97	2.87	4	-0.16
1039	0.062	1	0.068	-0.070	-5.52	18.00	4.34	5	-1.27
1040	0.068	1	0.072	-0.024	-4.37	18.00	4.34	5	-1.01
1041	0.600	0	0.476	0.000	0.50	13.80	2.30	4	0.22
1042	0.056	1	0.078	-0.239	-2.39	20.25	3.81	6	-0.63
1043	0.126	1	0.077	-1.306	-2.84	20.00	3.99	6	-0.71
1044	0.236	2	0.132	-1.915	-2.09	20.00	3.99	6	-0.52
1045	0.087	1	0.119	-0.266	-1.74	20.03	3.97	6	-0.44
1046	0.458	2	0.169	-2.028	-1.80	20.00	3.99	6	-0.45
1047	0.437	2	0.243	-0.543	-1.27	20.34	3.74	6	-0.34
1049	0.640	2	0.668	-0.037	0.57	17.20	3.97	5	0.14
1050	0.039	1	0.051	-0.283	-16.92	4.00	6.73	2	-2.51
1051	0.150	2	0.079	-0.256	-17.31	4.00	6.73	2	-2.57
1052	0.063	1	0.060	-0.015	-7.93	18.00	4.34	5	-1.83
1053	0.037	1	0.062	-0.824	-11.16	3.73	6.80	1	-1.64
1054	0.594	2	0.351	-8.774	-0.67	21.07	3.20	6	-0.21
1055	0.139	1	0.197	-0.964	-0.93	21.07	3.19	6	-0.29
1056	0.218	2	0.335	-1.623	-0.70	21.17	3.12	6	-0.23
1057	0.000	2	0.138	-5.199	-1.99	17.06	3.90	5	-0.51
1058	1.000	2	0.653	-5.112	0.40	17.29	4.01	5	0.10
1059	0.767	2	0.759	-0.003	2.12	19.14	4.14	5	0.51
1060	0.750	2	0.462	-0.859	-0.25	21.82	2.63	6	-0.09
1061	0.187	2	0.118	-0.142	-2.74	20.00	3.99	6	-0.69
1063	0.301	1	0.173	-3.678	-0.71	22.95	2.17	6	-0.33
1064	0.252	2	0.228	-0.113	-1.17	20.97	3.27	6	-0.36
1066	0.219	2	0.161	-0.284	-0.77	27.15	1.66	7	-0.47
1067	0.583	0	0.610	0.000	1.28	12.86	1.97	4	0.65
1068	0.625	0	0.610	0.000	1.28	12.86	1.97	4	0.65
1069	0.150	2	0.227	-0.374	-0.76	23.12	2.11	6	-0.36
1070	0.819	2	0.808	-0.022	1.50	23.61	1.94	6	0.77
1071	0.900	2	0.789	-1.575	1.33	23.46	1.99	6	0.67
1072	0.875	2	0.792	-0.494	1.70	22.00	2.50	6	0.68
1073	0.875	2	0.774	-0.232	1.20	23.29	2.05	6	0.59
1074	0.850	2	0.864	-0.153	3.06	21.87	2.59	6	1.18
1075	0.875	0	0.605	0.000	2.35	20.36	3.73	6	0.63
1076	0.938	2	0.824	-0.905	2.99	20.77	3.42	6	0.87
1077	0.889	2	0.800	-0.458	1.95	21.74	2.69	6	0.72
1078	0.781	2	0.787	-0.002	1.32	23.40	2.01	6	0.66
1079	0.833	2	0.800	-0.035	1.82	22.00	2.50	6	0.73
1080	0.750	2	0.782	-0.015	1.57	22.00	2.50	6	0.63
1081	0.125	1	0.118	-0.004	-0.71	27.28	1.62	7	-0.44
1082	0.250	2	0.181	-0.179	-0.88	26.15	2.04	7	-0.43
1083	0.625	2	0.661	-0.012	0.23	23.74	1.89	6	0.12

Appendix 3.3A continued.

Site	p(obs)	hab	Pexp	lnL	Dist	S	Width	Seg	Dist/W
1084	0.250	2	0.663	-3.209	0.25	23.62	1.94	6	0.13
1085	0.417	2	0.660	-1.112	0.23	23.54	1.96	6	0.12
1086	0.750	2	0.704	-0.026	0.55	23.41	2.01	6	0.27
1087	0.825	2	0.773	-0.238	1.19	23.27	2.06	6	0.58
1089	1.000	2	0.832	-1.652	3.32	20.59	3.55	6	0.93
1091	0.938	2	0.848	-1.203	4.12	20.08	3.94	6	1.05
1092	0.875	2	0.868	-0.001	4.88	20.00	3.99	6	1.22
1097	0.854	2	0.895	-0.309	5.97	20.00	3.99	6	1.50
1098	1.000	2	0.864	-1.320	4.71	20.00	3.99	6	1.18
1099	0.792	2	0.868	-2.978	4.88	20.00	3.99	6	1.22
1100	0.833	2	0.791	-0.209	1.05	11.07	1.56	3	0.67
1103	0.227	2	0.205	-0.048	-1.34	16.00	3.41	5	-0.39
1104	0.202	1	0.101	-1.769	-1.65	16.00	3.41	5	-0.48
1105	0.167	2	0.167	0.000	-1.55	16.00	3.41	5	-0.46
1109	0.000	2	0.183	-3.039	-1.43	20.86	3.35	6	-0.43
1110	0.577	2	0.290	-7.107	-0.95	16.00	3.41	5	-0.28
1111	1.000	2	0.965	-0.645	20.39	2.00	7.22	1	2.82
1112	0.854	2	0.961	-3.447	19.39	2.00	7.22	1	2.69
1113	0.609	2	0.524	-0.287	-0.10	15.49	3.14	4	-0.03
2012	0.464	2	0.290	-0.676	-0.95	16.00	3.41	5	-0.28
2054	0.125	2	0.351	-1.860	-0.67	21.07	3.20	6	-0.21
2115	0.093	2	0.124	-0.168	-2.17	20.00	3.99	6	-0.54
2116	0.081	1	0.072	-0.048	-4.37	18.00	4.34	5	-1.01
2117	0.074	2	0.151	-1.788	-1.65	16.00	3.41	5	-0.48
2118	0.813	2	0.872	-0.170	5.01	20.00	3.99	6	1.25
2119	0.071	2	0.101	-0.222	-2.36	27.12	1.67	7	-1.41
2120	0.114	2	0.100	-0.028	-2.51	27.05	1.70	7	-1.47
2121	0.020	2	0.110	-3.857	-1.75	26.95	1.74	7	-1.01
2122	0.938	2	0.902	-0.253	11.35	2.00	7.22	1	1.57
2124	0.813	2	0.899	-0.402	11.08	2.00	7.22	1	1.54
2126	0.828	2	0.898	-1.034	11.01	2.00	7.22	1	1.52
2127	0.969	2	0.899	-1.227	11.07	2.00	7.22	1	1.53
2132	0.938	2	0.897	-0.161	10.96	2.00	7.22	1	1.52
2133	0.922	2	0.899	-0.179	11.11	2.00	7.22	1	1.54
2134	0.886	2	0.898	-0.049	11.06	2.00	7.22	1	1.53
2135	0.095	2	0.111	-0.067	-1.68	26.93	1.75	7	-0.96
2136	0.900	2	0.899	0.000	11.09	2.00	7.22	1	1.54
2138	0.875	2	0.899	-0.021	11.12	2.00	7.22	1	1.54
2140	0.773	2	0.898	-1.846	11.02	2.00	7.22	1	1.53
2141	0.846	2	0.898	-0.269	11.02	2.00	7.22	1	1.53
2142	0.167	2	0.116	-0.126	-3.38	18.00	4.34	5	-0.78
2143	0.134	2	0.120	-0.083	-2.54	18.00	4.34	5	-0.58
2144	1.000	2	0.219	-6.071	-1.61	18.00	4.34	5	-0.37
2145	0.208	2	0.140	-0.345	-2.20	18.00	4.34	5	-0.51
2146	0.792	2	0.641	-1.721	0.36	17.72	4.21	5	0.09
2147	0.449	2	0.611	-1.391	0.23	18.58	4.24	5	0.05
2148	0.125	2	0.121	-0.001	-2.46	18.00	4.34	5	-0.57
2149	0.500	2	0.100	-1.526	-6.20	18.00	4.34	5	-1.43
2150	0.222	0	0.598	0.000	2.49	19.32	4.11	5	0.60
2151	0.097	2	0.197	-1.632	-0.93	22.56	2.30	6	-0.40
2152	0.094	1	0.082	-0.113	-2.35	18.00	4.34	5	-0.54

Appendix 3.3A continued.

Site	p(obs)	hab	Pexp	lnL	Dist	S	Width	Seg	Dist/W
2153	0.250	1	0.061	-1.652	-5.24	36.00	3.02	9	-1.73
2154	0.755	2	0.777	-0.076	1.17	35.24	1.94	9	0.60
2155	0.813	2	0.801	-0.009	2.22	36.00	3.02	9	0.73
2156	0.854	2	0.850	-0.006	3.21	36.00	3.02	9	1.06
2157	0.125	1	0.069	-0.275	-3.54	36.00	3.02	9	-1.17
2158	0.850	2	0.861	-0.014	3.48	36.00	3.02	9	1.15
2159	0.134	1	0.109	-0.156	-0.03	32.14	0.07	9	-0.46
2163	0.893	2	0.933	-0.519	14.66	2.00	7.22	1	2.03
2164	1.000	2	0.944	-2.013	6.82	36.00	3.02	9	2.26
2165	0.930	2	0.945	-0.264	6.91	36.00	3.02	9	2.29
2166	0.069	2	0.076	-0.059	-8.21	36.00	3.02	9	-2.72
2167	0.031	1	0.046	-0.107	-12.61	18.48	4.26	5	-2.96
2200	0.936	2	0.899	-0.615	11.07	2.00	7.22	1	1.53

Appendix 3.3B

The most likely cline with variable width and α fitted in one dimension. The total likelihood of this cline when F_{st} is 0.025 is $LogL_{126}$=-118.49 and when F_{st}=0.0068 then $LogL_{126}$=157.12. The total number of sites used to describe the cline is 134. These are sites with known habitat type.

Table legend:-

p(obs) is the observed gene frequency of the site ($=\bar{p}$)
H is the habitat type:-1=ponds, 1=puddles, 0=unknown
Dist/W is distance of a site from the centre of the cline measured in widths.
pexp is the expected frequency of a site when habitat is not taken into account.
pexp hab is the expected frequency of a site when habitat is taken into account such that $p=\alpha Hpq$.
Ne is the effective sample size allowing for a variance in gene frequency when F_{st}=0.025
LogL1; the likelihood of the model for a particular site when F_{st}=0.025; if >2 p<0.05
LogL2; the likelihood of the model for a particular site when F_{st}=0.0068; if >2 p<0.05
S is the distance along the cline from the starting point
Width is the width of the cline at distance S along the cline
Seg is the nearest segment to the site.

Site	H	p(obs)	Dist/w	pex	pexp hab	Ne F_{st}=0.025	Log L1	LogL2 (0.0068)
1	-1	0.079	-2.71	0.088	0.062	86	0.190	0.400
2	-1	0.067	-1.26	0.103	0.074	97	0.038	0.079
3	-1	0.059	-1.74	0.098	0.070	71	0.066	0.102
4	-1	0.068	-1.35	0.102	0.073	71	0.013	0.021
5	-1	0.121	-0.78	0.109	0.078	57	0.620	1.089
6	-1	0.121	-0.20	0.311	0.244	61	2.878	5.361
7	0	0.528	0.18	0.652	0.652	21		
8	1	0.800	0.35	0.718	0.782	39	0.040	0.053
9	1	0.800	0.56	0.746	0.805	24	0.002	0.003
10	-1	0.286	-0.69	0.110	0.079	16	2.958	3.647
11	-1	0.042	-0.84	0.108	0.078	22	0.240	0.266
12	1	0.650	0.84	0.779	0.833	16	1.606	1.880
13	0	0.604	0.14	0.615	0.615	17		
14	1	0.585	0.57	0.746	0.806	27	3.435	4.702
15	1	0.932	1.20	0.815	0.862	125	3.021	6.783
16	0	0.781	1.22	0.816	0.816	50		
17	1	0.918	1.42	0.833	0.877	101	0.865	1.611
18	1	0.885	1.15	0.810	0.859	81	0.244	0.409
19	1	0.919	1.76	0.860	0.898	111	0.282	0.572
20	1	0.955	1.59	0.847	0.888	71	2.037	2.909
1001	1	0.730	0.65	0.756	0.814	66	1.417	2.748
1002	1	0.684	0.15	0.622	0.696	56	0.020	0.038
1003	1	0.740	0.25	0.703	0.769	66	0.152	0.289
1004	-1	0.454	0.18	0.652	0.581	7	0.239	0.262
1005	-1	0.125	-0.13	0.368	0.295	4	0.366	0.380

Appendix 3.3B continued

Site	H	p(obs)	Dist/w	pex	pexp hab	Ne F_{st}=0.025	Log L1	LogL2 (0.0068)
1010	0	0.250	-0.09	0.406	0.406	3		
1011	1	0.381	-0.17	0.338	0.409	30	0.048	0.077
1013	0	0.181	-0.45	0.150	0.150	24		
1014	-1	0.023	-0.60	0.111	0.080	76	2.303	3.105
1016	1	0.042	-0.19	0.321	0.390	22	7.595	8.392
1018	0	0.500	-0.40	0.176	0.176	3		
1019	1	0.227	-0.49	0.134	0.170	13	0.139	0.160
1025	1	1.000	2.96	0.922	0.945	9	0.502	0.520
1028	1	0.983	2.96	0.922	0.945	55	1.055	1.268
1029	1	0.984	2.96	0.922	0.945	147	3.054	5.543
1032	0	0.000	-1.43	0.101	0.101	23		
1033	1	0.069	-0.97	0.107	0.136	59	1.349	1.841
1035	-1	0.080	-0.64	0.111	0.080	75	0.000	0.000
1036	-1	0.062	-0.66	0.110	0.079	13	0.027	0.029
1037	0	0.050	-0.20	0.309	0.309	18		
1038	1	0.367	-0.16	0.346	0.418	8	0.044	0.049
1039	-1	0.062	-1.27	0.103	0.074	105	0.120	0.260
1040	-1	0.068	-1.01	0.106	0.076	89	0.048	0.092
1041	0	0.600	0.22	0.699	0.699	4		
1042	-1	0.056	-0.63	0.111	0.080	47	0.205	0.265
1043	-1	0.126	-0.71	0.110	0.079	53	0.697	1.181
1044	1	0.236	-0.52	0.119	0.152	31	0.752	1.147
1045	-1	0.087	-0.44	0.157	0.116	39	0.162	0.214
1046	1	0.458	-0.45	0.151	0.191	8	1.470	1.624
1047	1	0.437	-0.34	0.212	0.265	5	0.378	0.405
1049	1	0.640	0.14	0.618	0.692	18	0.110	0.129
1050	-1	0.039	-2.51	0.089	0.064	107	0.641	0.978
1051	1	0.150	-2.57	0.089	0.114	9	0.049	0.053
1052	-1	0.063	-1.83	0.097	0.069	104	0.033	0.060
1053	-1	0.037	-1.64	0.099	0.071	83	0.877	1.318
1054	1	0.594	-0.21	0.304	0.371	42	4.308	7.362
1055	-1	0.139	-0.29	0.244	0.186	50	0.391	0.658
1056	1	0.218	-0.23	0.290	0.355	32	1.431	2.175
1057	1	0.000	-0.51	0.125	0.159	31	5.367	6.108
1058	1	1.000	0.10	0.577	0.654	12	5.010	5.190
1059	1	0.767	0.51	0.740	0.800	16	0.054	0.060
1060	1	0.750	-0.09	0.400	0.475	5	0.777	0.804
1061	1	0.187	-0.69	0.110	0.141	7	0.056	0.060
1063	-1	0.301	-0.33	0.218	0.164	39	2.247	4.260
1064	1	0.252	-0.36	0.200	0.251	40	0.000	0.000
1066	1	0.219	-0.47	0.144	0.183	20	0.081	0.102
1067	0	0.583	0.65	0.756	0.756	9		
1068	0	0.625	0.65	0.756	0.756	4		
1069	1	0.150	-0.36	0.200	0.250	17	0.502	0.587
1070	1	0.820	0.77	0.771	0.827	39	0.007	0.009
1071	1	0.900	0.67	0.759	0.816	30	0.823	0.963
1072	1	0.875	0.68	0.761	0.818	19	0.224	0.248
1073	1	0.875	0.59	0.749	0.808	7	0.109	0.113
1074	1	0.850	1.18	0.813	0.861	94	0.044	0.091
1075	0	0.875	0.63	0.754	0.754	13		

Appendix 3.3B continued

Site	H	p(obs)	Dist/w	pex	pexp hab	Ne F_{st}=0.025	Log L1	LogL2 (0.0068)
1076	1	0.938	0.87	0.782	0.836	15	0.711	0.761
1077	1	0.889	0.72	0.766	0.822	15	0.255	0.282
1078	1	0.781	0.66	0.757	0.815	19	0.069	0.078
1079	1	0.833	0.73	0.766	0.823	9	0.004	0.004
1080	1	0.750	0.63	0.754	0.812	5	0.059	0.061
1081	-1	0.125	-0.44	0.157	0.115	16	0.007	0.008
1082	1	0.250	-0.43	0.160	0.203	11	0.070	0.080
1083	1	0.625	0.12	0.598	0.673	4	0.019	0.020
1084	1	0.250	0.13	0.604	0.679	8	3.259	3.601
1085	1	0.417	0.12	0.595	0.671	8	1.052	1.162
1086	1	0.750	0.27	0.707	0.772	5	0.007	0.007
1087	1	0.825	0.58	0.748	0.807	25	0.025	0.030
1089	1	1.000	0.93	0.789	0.841	9	1.536	1.591
1091	1	0.938	1.05	0.800	0.850	28	1.017	1.158
1092	1	0.875	1.22	0.817	0.864	7	0.004	0.004
1097	1	0.854	1.50	0.840	0.882	31	0.110	0.132
1098	1	1.000	1.18	0.813	0.861	9	1.333	1.380
1099	1	0.792	1.22	0.817	0.864	72	1.405	2.569
1100	1	0.833	0.67	0.760	0.817	30	0.027	0.032
1103	1	0.227	-0.39	0.181	0.227	24	0.000	0.000
1104	-1	0.202	-0.48	0.136	0.099	28	1.335	1.871
1105	1	0.167	-0.46	0.149	0.189	10	0.017	0.019
1109	1	0.000	-0.43	0.163	0.206	14	3.300	3.418
1110	1	0.577	-0.28	0.252	0.311	28	4.218	5.915
1111	1	1.000	2.82	0.917	0.941	17	1.032	1.105
1112	1	0.854	2.69	0.911	0.937	31	1.337	1.607
1113	1	0.609	-0.03	0.456	0.534	17	0.192	0.230
2012	1	0.464	-0.28	0.252	0.311	9	0.468	0.525
2054	1	0.125	-0.21	0.304	0.371	12	1.880	2.077
2115	1	0.093	-0.54	0.112	0.143	29	0.332	0.406
2116	-1	0.081	-1.01	0.106	0.076	56	0.009	0.013
2117	1	0.075	-0.48	0.136	0.173	48	1.966	2.691
2118	1	0.813	1.25	0.820	0.866	11	0.124	0.133
2119	1	0.071	-1.41	0.101	0.130	34	0.605	0.746
2120	1	0.114	-1.47	0.101	0.129	22	0.024	0.028
2121	1	0.020	-1.01	0.106	0.136	53	4.473	5.411
2122	1	0.938	1.57	0.846	0.887	28	0.414	0.471
2124	1	0.813	1.54	0.843	0.885	11	0.243	0.260
2126	1	0.828	1.52	0.842	0.884	36	0.489	0.617
2127	1	0.969	1.53	0.843	0.884	29	1.394	1.587
2132	1	0.938	1.52	0.842	0.884	15	0.251	0.269
2133	1	0.922	1.54	0.843	0.885	46	0.343	0.433
2134	1	0.886	1.53	0.843	0.884	45	0.000	0.000
2135	1	0.095	-0.96	0.107	0.137	35	0.291	0.378
2136	1	0.900	1.54	0.843	0.885	17	0.020	0.022
2138	1	0.875	1.54	0.843	0.885	7	0.003	0.003
2140	1	0.773	1.53	0.842	0.884	24	1.162	1.378
2141	1	0.846	1.53	0.842	0.884	19	0.119	0.132
2142	1	0.167	-0.78	0.109	0.139	10	0.031	0.034
2143	1	0.134	-0.58	0.111	0.142	55	0.015	0.028

Appendix 3.3B continued

Site	H	p(obs)	Dist/w	pex	pexp hab	Ne F_{st}=0.025	Log L1	LogL2 (0.0068)
2144	1	1.000	-0.37	0.193	0.242	4	5.096	5.458
2145	1	0.208	-0.51	0.126	0.161	16	0.125	0.150
2146	1	0.792	0.09	0.565	0.643	27	1.403	1.687
2147	1	0.449	0.05	0.536	0.615	20	1.120	1.470
2148	1	0.125	-0.57	0.111	0.143	16	0.021	0.024
2149	1	0.500	-1.43	0.101	0.130	3	1.211	1.254
2150	0	0.222	0.60	0.751	0.751	13		
2151	1	0.097	-0.40	0.174	0.220	34	1.784	2.305
2152	-1	0.094	-0.54	0.112	0.081	64	0.079	0.135
2153	-1	0.250	-1.73	0.098	0.070	8	1.323	1.462
2154	1	0.755	0.60	0.751	0.810	40	0.363	0.502
2155	1	0.813	0.73	0.767	0.823	20	0.008	0.009
2156	1	0.854	1.06	0.802	0.852	51	0.001	0.002
2157	-1	0.125	-1.17	0.104	0.075	12	0.190	0.210
2158	1	0.850	1.15	0.810	0.859	27	0.008	0.009
2159	-1	0.134	-0.46	0.144	0.106	37	0.147	0.210
2163	1	0.893	2.03	0.877	0.911	39	0.073	0.090
2164	1	1.000	2.26	0.890	0.921	31	2.562	2.916
2165	1	0.929	2.29	0.892	0.922	77	0.031	0.047
2166	1	0.069	-2.72	0.087	0.113	90	0.964	1.895
2167	-1	0.031	-2.96	0.085	0.061	32	0.293	0.333
2200	1	0.936	1.53	0.843	0.884	54	0.818	1.075

APPENDICES For Chapter 5

APPENDIX 5.1 - Egg sizes

Egg diameters (mm) of five eggs from some of the egg batches collected for the translocation experiment. Sites 2039 and 2116 are ponds and the remaining sites are puddles. The mean diameter (mm) and mean egg volume (mm^3) are given for each batch. * indicates missing data. See Chapter 5 for details.

Site	Batch	Egg 1	Egg 2	Egg 3	Egg 4	Egg 5	Mean diam	Vol
2039	1	1.60	1.49	1.63	1.52	1.49	1.55	1.93
2039	2	1.55	1.60	1.60	1.58	1.58	1.58	2.07
2039	3	1.40	1.49	1.55	1.46	1.40	1.46	1.62
2039	4	1.31	1.52	1.52	1.52	1.58	1.49	1.72
2039	5	1.58	1.49	1.52	1.58	1.58	1.55	1.93
2039	6	1.55	1.58	1.55	1.37	1.58	1.52	1.85
2039	7	1.40	1.40	1.34	1.37	1.37	1.38	1.37
2039	8	1.55	1.63	1.58	1.58	1.58	1.58	2.07
2039	9	1.52	1.58	1.60	1.63	1.60	1.59	2.09
2039	10	1.55	1.52	1.55	1.58	1.58	1.55	1.96
2039	11	1.58	1.52	1.53	1.63	1.58	1.57	2.01
2039	12	1.52	1.58	1.60	1.63	1.49	1.56	2.00
2039	13	1.49	1.37	1.55	1.52	1.55	1.49	1.74
2039	14	1.43	1.60	1.43	1.43	1.52	1.48	1.70
2039	15	1.60	1.55	1.49	1.52	1.40	1.51	1.80
2039	16	1.46	1.43	1.60	1.46	1.66	1.52	1.85
2039	17	1.52	1.55	1.14	1.25	1.58	1.41	1.45
2039	18	1.60	1.46	1.46	1.46	1.60	1.52	1.83
2039	19	1.49	1.52	1.46	1.52	1.46	1.49	1.72
2116	1	1.46	1.55	1.43	1.49	1.43	1.47	1.66
2116	2	1.44	1.40	1.40	1.31	1.46	1.40	1.45
2116	3	1.75	1.60	1.60	1.60	1.66	1.64	2.33
2116	4	1.40	1.34	1.34	1.31	1.17	1.31	1.18
2116	5	1.46	1.43	1.43	1.49	1.46	1.45	1.60
2116	6	1.49	1.52	1.58	1.31	1.43	1.46	1.64
2116	7	1.49	1.49	1.49	1.46	1.49	1.48	1.70
2116	8	1.37	1.31	1.37	1.31	1.37	1.35	1.28
2116	9	1.55	1.46	1.46	1.55	1.52	1.50	1.78
2116	10	1.46	1.52	1.52	1.58	1.52	1.52	1.83
2116	11	1.40	1.34	0.20	1.36	1.34	1.13	0.75
2116	12	1.37	1.37	1.34	1.34	1.34	1.35	1.30
2116	13	1.40	1.40	1.34	1.25	1.46	1.37	1.35
2116	14	1.43	1.40	1.34	1.25	1.37	1.36	1.31
2116	15	1.58	1.55	1.52	1.58	*	1.55	1.96
2122	1	2.36	2.36	2.39	2.39	2.42	2.39	7.11
2122	2	2.36	*	*	*	*	2.36	6.90
2122	3	2.36	2.51	2.48	2.33	2.42	2.42	7.43
2122	4	2.10	1.98	2.04	2.04	*	2.04	4.45
2122	5	2.51	2.51	*	*	*	2.51	8.26

Appendix 5.1 continued

Site	Batch	Egg 1	Egg 2	Egg 3	Egg 4	Egg 5	Mean diam	Vol
2122	6	2.42	2.33	2.54	2.51	2.51	2.46	7.81
2122	7	2.60	2.60	2.54	2.28	2.36	2.47	7.92
2122	8	2.45	2.51	2.42	2.45	2.42	2.45	7.70
2122	9	2.48	2.42	2.36	2.33	2.54	2.43	7.48
2122	10	2.48	2.45	2.45	2.48	2.48	2.47	7.86
2122	11	2.57	2.54	2.51	2.42	2.51	2.51	8.26
2126	1	2.48	2.33	2.39	2.39	2.42	2.40	7.27
2126	2	2.07	2.04	2.04	2.01	2.10	2.05	4.53
2126	3	2.45	2.42	2.36	2.39	2.42	2.41	7.32
2126	4	2.48	2.36	2.54	2.48	2.45	2.46	7.81
2127	1	2.71	2.77	2.89	2.77	2.74	2.78	11.20
2127	2	2.48	2.42	2.48	2.45	2.42	2.45	7.70
2127	3	2.54	2.39	2.30	2.33	2.51	2.41	7.37
2127	4	2.54	2.39	2.57	2.22	2.28	2.40	7.21
2127	5	2.80	2.86	2.80	2.89	2.80	2.83	11.85
2127	6	2.45	2.33	2.51	2.36	2.42	2.41	7.37
2127	7	2.74	2.68	2.89	2.54	2.80	2.73	10.65
2127	8	2.77	2.74	2.77	2.74	2.77	2.76	10.99
2131	1	2.36	2.48	2.42	2.39	2.54	2.44	7.59
2131	2	2.54	2.45	2.48	2.39	2.30	2.43	7.53
2131	3	2.30	2.32	2.36	2.28	2.22	2.30	6.33
2131	4	2.39	2.25	2.36	*	*	2.33	6.65
2131	5	2.42	2.39	2.39	2.45	2.51	2.43	7.53
2131	6	2.39	2.39	2.30	2.42	2.33	2.37	6.95
2131	7	2.51	2.45	2.45	2.33	2.36	2.42	7.43
2131	8	2.51	2.48	2.39	2.60	2.65	2.53	8.43
2131	9	2.42	2.62	2.57	2.45	2.51	2.51	8.32
2131	10	2.48	2.51	2.51	2.54	2.33	2.47	7.92
2131	11	2.22	2.25	2.22	2.39	2.33	2.28	6.21
2131	12	2.33	2.36	2.33	2.33	2.33	2.34	6.70
2131	13	2.36	2.36	2.30	2.33	2.36	2.35	6.75
2131	14	2.39	2.51	2.42	2.39	2.45	2.43	7.53
2131	15	2.62	2.33	2.62	2.51	2.62	2.54	8.61
2131	16	2.39	2.33	2.33	2.33	*	2.35	6.78
2132	1	2.54	2.45	2.42	2.51	2.51	2.48	8.03
2132	2	2.83	2.86	2.83	2.86	2.83	2.84	12.00
2132	3	2.51	2.48	2.48	2.42	2.39	2.46	7.75
2132	4	2.57	2.39	2.51	2.39	2.48	2.47	7.86
2134	1	2.16	2.16	2.13	2.13	2.19	2.15	5.22
2134	2	2.54	2.48	2.51	2.42	2.54	2.50	8.15
2134	3	2.48	2.36	2.42	2.54	2.36	2.43	7.53
2134	4	2.51	2.48	2.39	2.39	2.42	2.44	7.59
2134	5	2.45	2.42	2.48	2.42	2.39	2.43	7.53
2134	6	2.51	2.28	2.54	2.48	2.51	2.46	7.81
2134	7	2.19	2.22	2.19	2.22	2.19	2.20	5.57
2134	8	2.45	2.42	2.45	2.33	2.42	2.41	7.37
2134	10	2.57	2.48	2.45	2.51	2.33	2.47	7.86
2134	11	2.51	2.60	2.45	*	*	2.52	8.36
2134	12	2.48	2.54	2.45	2.42	2.33	2.44	7.64
2134	13	2.60	2.97	2.83	2.92	2.86	2.84	11.93

APPENDIX 5.2

The colour, weight, length and stage of each individual removed from an enclosure.

legend

Enc: the name of the enclosure; 'p' is a puddle enclosure and V is a pond enclosure.
Ind: individual name
colour: the colour code of the individual 1=orange, 2=grey, 0=unknown.
Stage: see Chapter 5 for details.
*: missing data

Enc	Ind	colour	weight mg	length mm	stage 1-8
p1	1	2	93	8.0	0
p1	2	2	106	8.0	0
p1	3	2	55	7.0	0
p1	4	2	90	8.0	0
p1	5	2	153	10.0	0
p1	6	2	102	8.0	0
p1	7	2	115	9.0	0
p1	8	2	69	8.0	0
p1	9	2	150	10.0	0
p1	10	2	77	8.0	0
p1	11	2	105	9.0	0
p1	12	2	105	8.0	0
p1	13	2	144	9.0	0
p1	14	2	52	6.0	0
p1	15	2	96	8.0	0
p1	16	2	91	8.0	0
p1	17	2	92	8.0	0
p1	18	2	77	7.0	0
p1	19	2	140	10.0	0
p1	20	2	66	8.0	0
p1	21	2	120	9.0	0
p1	22	2	213	11.0	0
p1	23	2	150	10.0	0
p1	24	2	86	8.0	0
p1	25	2	74	8.0	0
p1	26	2	88	8.0	0
p1	27	2	73	7.0	0
p1	28	2	134	9.0	0
p1	29	2	52	7.0	0
p1	30	2	77	8.0	0
p1	31	2	85	8.0	0
p1	32	2	85	7.0	0
p1	33	2	108	8.0	0
p1	34	2	111	9.0	0
p1	35	2	90	8.0	0
p1	36	2	70	8.0	0
p1	37	2	74	7.0	0
p1	38	2	80	8.0	0
p1	39	2	109	9.0	0
p1	40	2	64	7.0	0
p1	41	1	8	*	
p2	1	2	45	7.0	0
p2	2	2	137	10.0	0
p2	3	2	105	9.0	0
p2	4	2	121	10.0	0
p2	5	2	72	8.0	0
p2	6	2	89	8.0	0
p2	7	2	61	7.0	0
p2	8	2	70	8.0	0
p2	9	2	66	8.0	0
p2	10	2	69	7.0	0
p2	11	2	76	8.0	0
p2	12	2	165	11.0	0
p2	13	2	134	9.0	0
p2	14	2	271	12.0	0
p2	15	2	81	9.0	0
p2	16	2	106	9.0	0
p2	17	2	176	10.0	0
p2	18	2	75	8.0	0
p2	19	2	140	9.0	0
p2	20	2	114	8.0	0
p2	21	2	60	7.0	0
p2	22	2	117	9.0	0
p2	23	2	157	10.0	0
p2	24	2	106	9.0	0
p2	25	2	112	10.0	0
p2	26	2	291	12.0	0
p2	27	2	119	9.0	0
p2	28	2	172	10.0	0
p2	29	2	137	9.0	0
p2	30	2	210	10.0	0
p2	31	2	109	9.0	0
p2	32	2	93	8.0	0
p2	33	2	96	9.0	0
p2	34	2	97	8.0	0
p2	35	2	125	9.0	0
p2	36	2	139	10.0	0
p2	37	2	109	8.0	0
p2	38	2	100	8.0	0
p2	39	2	121	10.0	0
p2	40	2	96	8.0	0
p2	41	2	79	8.0	0
p2	42	2	117	9.0	0
p2	43	2	145	9.0	0
p2	44	2	113	8.0	0
p2	45	2	74	8.0	0
p2	46	2	87	8.0	0

Appendix 5.2 continued (2)

Enc	Ind	colour	weight mg	length mm	stage 1-8
p2	47	2	100	8.0	0
p2	48	2	65	7.0	0
p2	49	2	73	8.0	0
p2	50	2	100	8.0	0
p2	51	2	71	8.0	0
p2	52	2	78	8.0	0
p2	53	2	120	9.0	0
p2	54	2	41	*	0
p2	55	1	10	*	0
p2	56	1	21	*	0
p2	57	1	12	*	0
p2	58	1	11	*	0
p2	59	1	5	*	0
p2	60	1	5	*	0
p2	61	1	7	*	0
p3	1	2	192	9.0	0
p3	2	2	117	9.0	0
p3	3	2	129	9.0	0
p3	4	2	106	9.0	0
p3	5	2	64	7.0	0
p3	6	2	130	9.0	0
p3	7	2	133	9.0	0
p3	8	2	77	8.0	0
p3	9	2	165	10.0	0
p3	10	2	105	9.0	0
p3	11	2	130	8.0	0
p3	12	2	168	10.0	0
p3	13	2	139	9.0	0
p3	14	2	151	9.0	0
p3	15	2	153	10.0	0
p3	16	2	139	9.0	0
p3	17	2	97	8.0	0
p3	18	2	230	11.0	0
p3	19	2	203	11.0	0
p3	20	2	78	8.0	0
p3	21	2	154	10.0	0
p3	22	1	19	*	0
p3	23	1	21	*	0
p3	24	1	12	*	0
p3	25	2	21	*	0
p3	26	2	52	*	0
p3	27	1	2	*	0
p4	1	2	287	12.0	0
p4	2	2	243	12.0	0
p4	3	2	110	9.0	0
p4	4	2	121	9.0	0
p4	5	2	105	8.0	0
p4	6	2	72	7.0	0
p4	7	2	110	9.0	0
p4	8	2	107	9.0	0
p4	9	2	418	13.0	2
p4	10	2	137	9.0	0
p4	11	2	115	9.0	0
p4	12	2	350	13.0	0
p4	13	2	344	12.0	0
p4	14	2	122	10.0	0
p4	15	2	111	9.0	0
p4	16	2	62	7.0	0
p4	17	2	91	8.0	0
p4	18	2	144	10.0	0
p4	19	2	133	9.0	0
p4	20	2	202	11.0	0
p4	21	2	110	9.0	0
p4	22	2	207	10.0	0
p4	23	2	95	8.0	0
p4	24	2	109	9.0	0
p4	25	2	110	8.0	0
p4	26	2	120	8.0	0
p4	27	2	134	9.0	0
p4	28	2	89	7.0	0
p4	29	2	211	10.0	0
p4	30	2	120	9.0	0
p4	31	2	102	9.0	0
p4	32	2	63	8.0	0
p4	33	2	82	7.0	0
p4	34	2	124	8.0	0
p4	35	2	96	8.0	0
p4	36	2	76	7.0	0
p4	37	1	9	*	0
p4	38	1	10	*	0
p4	39	2	84	*	0
p4	40	1	4	*	0
p4	41	1	13	*	0
p4	42	1	5	*	0
p4	43	1	10	*	0
p6	1	2	179	10.0	0
p6	2	2	169	10.0	0
p6	3	2	219	10.0	0
p6	4	2	135	9.0	0
p6	5	2	96	8.0	0
p6	6	2	228	11.0	0
p6	7	2	169	10.0	0
p6	8	2	135	10.0	0
p6	9	2	148	9.0	0
p6	10	2	108	9.0	0
p6	11	2	72	8.0	0
p6	12	2	129	9.0	0
p6	13	2	219	11.0	0
p6	14	2	204	10.0	0
p6	15	2	119	8.0	0
p6	16	2	82	8.0	0
p6	17	2	175	10.0	0
p6	18	2	151	8.0	0
p6	19	2	103	8.0	0
p6	20	2	160	10.0	0

Appendix 5.2 continued (3)

Enc	Ind	colour	weight mg	length mm	stage 1-8
p6	21	2	229	12.0	0
p6	22	2	188	11.0	0
p6	23	2	138	10.0	0
p6	24	2	147	9.0	0
p6	25	2	177	10.0	0
p6	26	2	98	9.0	0
p6	27	2	305	12.0	0
p6	28	2	181	10.0	0
p6	29	2	106	9.0	0
p6	30	2	131	9.0	0
p6	31	2	218	10.0	0
p6	32	2	183	10.0	0
p6	33	2	96	9.0	0
p6	34	2	160	10.0	0
p6	35	2	99	8.0	0
p6	36	2	184	10.0	0
p6	37	2	114	9.0	0
p6	38	2	99	8.0	0
p6	39	2	104	8.0	0
p6	40	2	133	9.0	0
p6	41	2	123	9.0	0
p6	42	2	72	7.0	0
p6	43	2	90	8.0	0
p6	44	2	140	10.0	0
p6	45	2	113	8.0	0
p6	46	2	108	8.0	0
p6	47	2	77	8.0	0
p6	48	2	113	9.0	0
p6	49	2	273	11.0	0
p6	50	2	140	9.0	0
p6	51	2	167	10.0	0
p6	52	2	117	9.0	0
p6	53	2	123	8.0	0
p6	54	2	93	8.0	0
p6	55	2	164	9.0	0
p6	56	1	16	*	0
p6	57	1	16	*	0
p6	58	1	20	*	0
p6	59	1	16	*	0
p6	60	1	13	*	0
p6	61	1	9	*	0
p6	62	1	17	*	0
p6	63	1	9	*	0
p6	64	1	10	*	0
p6	65	1	14	*	0
p6	66	1	7	*	0
p6	67	1	22	*	0
p6	68	1	5	*	0
p6	69	1	22	*	0
p6	70	1	28	*	0
v1	1	0	175	7.0	0
v1	2	0	543	14.0	0
v1	3	0	607	13.0	4
v1	4	0	248	7.0	0
v1	5	0	303	9.0	0
v1	6	0	189	7.0	0
v1	7	0	92	5.0	0
v1	8	0	297	9.0	0
v1	9	0	283	8.0	0
v1	10	0	322	8.0	0
v1	11	0	263	8.0	0
v1	12	0	496	11.0	0
v1	13	0	237	8.0	0
v1	14	0	344	10.0	0
v1	15	0	232	8.0	0
v1	16	0	250	8.0	0
v1	17	0	256	9.0	0
v1	18	0	608	13.5	0
v1	19	0	281	8.0	0
v1	20	0	463	12.0	0
v1	21	0	702	14.0	2
v1	22	0	493	11.0	2
v1	23	0	212	7.0	0
v1	24	0	205	7.0	0
v1	25	0	602	14.0	4
v1	26	0	525	14.0	4
v1	27	0	236	8.0	0
v1	28	0	225	7.0	0
v1	29	0	174	7.0	0
v1	30	0	291	9.0	0
v1	31	0	150	7.0	0
v1	32	0	625	13.0	4
v1	33	0	152	10.0	0
v1	34	0	658	15.0	4
v1	35	0	373	13.5	4
v1	36	0	218	11.0	0
v1	37	0	272	12.0	0
v1	38	0	206	10.0	0
v1	39	0	314	13.0	4
v1	40	0	401	14.0	4
v1	41	0	530	14.0	6
v1	42	0	483	15.0	4
v1	43	0	580	15.0	4
v1	44	0	389	14.0	4
v1	45	0	407	13.0	4
v1	46	0	599	15.0	4
v1	47	0	577	16.0	6
v1	48	0	407	15.0	8
v1	49	0	424	14.0	4
v1	50	0	470	14.0	4
v1	51	0	460	14.0	4
v1	52	0	151	9.0	0
v1	53	0	573	15.0	4
v1	54	0	288	9.0	0
v1	55	0	467	15.0	4

Appendix 5.2 continued (4)

Enc	Ind	colour	weight mg	length mm	stage 1-8
v1	56	0	557	16.0	6
v1	57	0	457	14.0	4
v1	58	0	415	14.0	4
v1	59	0	231	11.0	0
v1	60	0	447	14.0	4
v1	61	0	91	9.0	0
v1	62	0	541	15.0	4
v1	63	0	509	16.0	4
v1	64	0	407	14.0	4
v1	65	0	633	16.0	6
v1	66	0	309	13.0	4
v1	67	0	595	17.0	4
v1	68	0	253	12.0	4
v1	69	0	769	17.0	6
v1	70	0	503	15.0	4
v1	71	0	574	16.0	6
v1	72	0	290	12.0	4
v1	73	0	373	14.0	4
v1	74	0	384	14.0	4
v1	75	0	385	13.0	4
v1	76	0	636	16.0	6
v1	77	0	520	15.0	4
v1	78	0	310	12.0	4
v2	1	1	240	11.0	0
v2	2	1	270	11.0	2
v2	3	1	292	12.0	2
v2	4	1	255	10.0	2
v2	5	1	222	10.0	2
v2	6	1	300	12.0	2
v2	7	1	238	10.0	2
v2	8	1	311	12.0	2
v2	9	2	526	15.0	2
v2	10	1	299	11.0	2
v2	11	1	156	9.0	0
v2	12	1	172	10.0	0
v2	13	1	90	8.0	0
v2	14	1	176	10.0	0
v2	15	1	184	10.0	0
v2	16	1	142	9.0	0
v2	17	2	559	16.0	2
v2	18	2	586	17.0	4
v2	19	2	635	16.0	2
v2	20	1	347	12.0	4
v2	21	1	228	11.0	2
v2	22	2	584	15.0	2
v2	23	2	680	17.0	4
v2	24	2	561	15.0	2
v2	25	2	759	17.0	4
v2	26	1	173	10.0	0
v2	27	1	362	13.0	2
v3	1	2	374	13.0	2
v3	2	1	294	12.0	2
v3	3	1	214	11.0	2
v3	4	1	272	11.0	2
v3	5	1	274	11.0	2
v3	6	1	179	10.0	0
v3	7	1	245	10.0	2
v3	8	1	373	12.0	2
v3	9	2	287	13.0	0
v3	10	1	246	11.0	2
v3	11	2	414	14.0	2
v3	12	1	261	11.0	2
v3	13	1	157	11.0	2
v3	14	1	335	12.0	2
v3	15	2	560	15.0	4
v3	16	2	682	16.0	6
v3	17	1	284	12.0	2
v3	18	1	241	11.0	2
v3	19	1	209	10.0	0
v3	20	2	656	16.0	4
v3	21	2	467	15.0	2
v3	22	2	462	15.0	6
v3	23	2	400	13.0	2
v3	24	2	366	14.0	2
v3	25	2	413	14.0	2
v3	26	2	353	14.0	2
v3	27	1	188	10.0	0
v3	28	2	520	15.0	2
v3	29	2	614	15.0	2
v3	30	2	462	14.0	2
v3	31	2	601	17.0	2
v3	32	2	532	16.0	2
v3	33	2	426	14.0	2
v3	34	2	597	16.0	2
v3	35	2	316	12.0	0
v3	36	2	371	12.0	0
v3	37	2	493	14.0	2
v3	38	2	182	11.0	0
v3	39	2	531	16.0	7
v3	40	2	381	13.0	2
v3	41	2	527	16.0	2
v3	42	2	228	12.0	0
v3	43	2	573	15.0	2
v3	44	2	380	13.0	2
v3	45	2	219	10.0	0
v3	46	2	457	14.0	2
v3	47	2	652	17.0	6
v3	48	2	552	14.0	2
v3	49	2	602	15.0	6
v3	50	1	401	14.0	2
v3	51	1	103	8.0	0
v4	1	0	213	11.0	0
v4	2	0	250	11.0	0
v4	3	0	600	16.0	6
v4	4	0	263	12.0	0

Appendix 5.2 continued (5)

Enc	Ind	colour	weight mg	length mm	stage 1-8
v4	5	0	230	11.0	0
v4	6	0	507	15.0	2
v4	7	0	513	14.0	6
v4	8	0	365	13.0	2
v4	9	0	437	14.0	2
v4	10	0	275	11.0	2
v4	11	0	223	12.0	0
v4	12	0	317	13.0	2
v4	13	0	297	11.0	2
v4	14	0	274	12.0	2
v4	15	0	171	10.0	0
v4	16	0	506	14.0	2
v4	17	1	345	12.0	2
v4	18	2	481	15.0	6
v4	19	2	567	15.0	4
v4	20	1	199	10.0	0
v4	21	2	300	12.0	2
v4	22	2	524	14.0	2
v4	23	2	533	15.0	2
v4	24	2	503	15.0	2
v4	25	2	490	15.0	2
v4	26	1	270	12.0	2
v4	27	2	581	16.0	2
v4	28	1	189	10.0	2
v4	29	1	294	12.0	2
v4	30	2	606	15.0	6
v4	31	2	375	13.0	2
v4	32	2	566	16.0	6
v4	33	1	273	11.0	2
v4	34	1	226	11.0	2
v4	35	2	389	12.0	2
v4	36	2	511	15.0	4
v4	37	1	293	12.0	2
v4	38	2	617	15.0	6
v4	39	1	224	10.0	2
v4	40	1	226	11.0	2
v4	41	2	557	16.0	2
v4	42	2	411	15.0	2
v4	43	2	467	14.0	2
v4	44	2	700	15.0	4
v4	45	2	467	14.0	2
v4	46	2	608	16.0	2
v4	47	2	277	12.0	2
v4	48	2	491	15.0	2
v4	49	2	558	16.0	4
v4	50	2	250	12.0	2
v4	51	1	148	9.0	2
v4	52	2	569	15.0	4
v4	53	2	429	14.0	2
v4	54	1	200	10.0	2
v4	55	2	542	15.0	4
v4	56	2	402	14.0	2
v4	57	2	599	16.0	2
v4	58	2	449	15.0	2
v4	59	2	431	15.0	8
v5	1	0	554	15.0	2
v5	2	0	635	14.0	2
v5	3	0	581	16.0	2
v5	4	0	610	14.0	4
v5	5	0	599	14.0	4
v5	6	0	502	16.0	2
v5	7	0	444	15.0	2
v5	8	0	471	14.0	2
v5	9	0	679	16.0	4
v5	10	0	602	16.0	2
v5	11	0	509	16.0	2
v5	12	0	643	16.0	2
v5	13	0	507	15.0	2
v5	14	0	515	16.0	2
v5	15	0	383	13.0	2
v6	1	1	169	10.0	2
v6	2	2	475	14.0	2
v6	3	2	478	15.0	2
v6	4	2	440	13.0	2
v6	5	2	393	14.0	2
v6	6	2	543	16.0	4
v6	7	2	418	14.0	2
v6	8	2	580	15.0	6
v6	9	2	591	16.0	6
v6	10	2	580	16.0	2
v6	11	2	474	15.0	2
v6	12	1	161	10.0	2
v6	13	2	432	14.0	2
v6	14	2	539	15.0	2
v6	15	2	445	14.0	2
v6	16	2	376	14.0	2
v6	17	1	160	10.0	2
v6	18	2	389	13.0	2
v6	19	2	493	15.0	4
v6	20	2	547	16.0	2
v6	21	2	678	16.0	6
v6	22	2	557	14.0	4
v6	23	2	394	14.0	2
v6	24	2	466	14.0	2
v6	25	1	102	8.0	0
v6	26	2	516	15.0	2
v6	27	2	652	16.0	4
v6	28	2	445	15.0	2
v6	29	2	588	16.0	2
v6	30	2	482	15.0	4
v6	31	2	449	14.0	2
v6	32	2	469	15.0	2
v6	33	2	388	14.0	2
v6	34	2	419	15.0	2
v6	35	2	374	13.0	2

Appendix 5.2 continued (6)

Enc	Ind	colour	weight mg	length mm	stage 1-8
v6	36	2	551	15.0	2
v6	37	2	411	13.0	2
v6	38	2	522	15.0	6
v6	39	2	420	13.0	2
v6	40	2	534	14.0	4
v6	41	2	506	15.0	6
v6	42	2	462	14.0	2
v6	43	2	462	13.0	2
v6	44	2	532	14.0	6
v6	45	2	408	13.0	2
v6	46	2	446	13.0	6
v6	47	2	530	16.0	2
v6	48	2	438	14.0	2
v6	49	2	419	14.0	8
v7	1	0	152	10.0	0
v7	2	0	112	8.0	0
v7	3	0	104	8.0	0
v7	4	0	194	10.0	0
v7	5	0	589	16.0	4
v7	6	0	568	15.0	4
v7	7	0	217	11.0	0
v7	8	0	99	8.0	0
v7	9	0	181	11.0	0
v7	10	0	188	10.0	0
v7	11	0	129	9.0	0
v7	12	0	191	10.0	0
v7	13	0	126	9.0	0
v7	14	0	187	10.0	0
v7	15	0	127	9.0	0
v7	16	0	617	17.0	4
v7	17	0	647	16.0	4
v7	18	0	189	10.0	0
v7	19	0	577	15.0	4
v7	20	0	558	15.0	2
v7	21	0	113	9.0	0
v7	22	0	149	11.0	0
v7	23	0	673	15.0	2
v7	24	0	518	14.0	2
v7	25	0	680	16.0	2
v7	26	0	563	14.0	2
v7	27	0	185	10.0	0
v7	28	0	114	8.0	0
v7	29	0	120	8.0	0
v7	30	0	622	15.0	4
v7	31	0	623	16.0	2
v7	32	0	563	15.0	2
v7	33	0	628	16.0	2
v7	34	0	425	14.0	2
v7	35	0	559	16.0	2
v7	36	0	433	14.0	2
v7	37	0	519	14.0	2
v7	38	0	602	16.0	2
v7	39	0	576	15.0	2
v7	40	0	556	16.0	2
v7	41	0	538	16.0	2
v7	42	0	108	9.0	0
v7	43	0	341	13.0	2
v7	44	0	451	14.0	2
v7	45	0	599	15.0	2
v7	46	0	653	16.0	2
v7	47	0	433	14.0	2
v7	48	0	608	15.0	2
v7	49	0	567	14.0	2
v7	50	0	410	14.0	2
v7	51	0	584	16.0	2
v7	52	0	536	15.0	2
v7	53	0	422	15.0	2
v7	54	0	452	15.0	2
v7	55	0	494	15.0	2
v7	56	0	305	13.0	2
v7	57	0	513	15.0	2
v7	58	0	711	17.0	6
v7	59	0	638	16.0	6
v7	60	0	633	15.0	2
v7	61	0	628	16.0	4
v7	62	0	528	15.0	2
v7	63	0	504	15.0	2
v7	64	0	451	14.0	2
v7	65	0	559	16.0	2
v7	66	0	163	10.0	0
v7	67	0	455	15.0	2
v7	68	0	706	16.0	4
v7	69	0	596	16.0	2
v7	70	0	646	16.0	4
v7	71	0	675	16.0	6
v8	1	1	81	8.0	0
v8	2	2	221	11.0	2
v8	3	1	70	8.0	0
v8	4	2	241	11.0	2
v8	5	2	338	13.0	2
v8	6	2	315	13.0	2
v8	7	1	359	13.0	2
v8	8	2	292	13.0	2
v8	9	2	366	13.0	2
v8	10	2	382	13.0	2
v8	11	2	292	13.0	2
v8	12	1	79	8.0	0
v8	13	2	249	11.0	2
v8	14	2	283	11.0	2
v8	15	2	299	12.0	2
v8	16	2	283	12.0	2
v8	17	2	279	12.0	2
v8	18	2	331	12.0	2
v8	19	2	306	14.0	2
v8	20	2	328	13.0	2

Appendix 5.2 continued (7)

Enc	Ind	colour	weight mg	length mm	stage 1-8
v8	21	2	329	13.0	2
v8	22	2	276	12.0	2
v8	23	2	299	13.0	2
v8	24	2	71	8.0	0
v8	25	2	236	11.0	2
v8	26	2	241	12.0	2
v8	27	2	355	12.0	2
v8	28	2	278	14.0	2
v8	29	2	150	9.0	0
v8	30	2	246	12.0	2
v8	31	2	349	12.0	2
v8	32	2	255	12.0	2
v8	33	2	225	11.0	0
v8	34	2	246	12.0	2
v8	35	2	438	14.0	2
v8	36	2	247	12.0	2
v8	37	2	461	14.0	2
v8	38	2	264	12.0	2
v8	39	2	199	10.0	0
v8	40	2	271	11.0	0
v8	41	2	195	11.0	2
v8	42	2	249	11.0	2
v8	43	2	282	12.0	0
v8	44	1	42	*	0
v9	1	2	230	11.0	2
v9	2	2	247	11.0	2
v9	3	2	304	12.0	0
v9	4	2	309	13.0	0
v9	5	2	205	10.0	0
v9	6	2	256	11.0	0
v9	7	2	212	11.0	0
v9	8	2	307	13.0	0
v9	9	2	180	10.0	0
v9	10	2	159	10.0	0
v9	11	1	85	7.0	0
v9	12	1	82	7.0	0
v9	13	2	238	12.0	0
v9	14	2	275	13.0	0
v9	15	2	242	11.0	2
v9	16	2	260	13.0	0
v9	17	2	235	12.0	0
v9	18	2	284	12.0	0
v9	19	2	195	10.0	0
v9	20	2	198	10.0	0
v9	21	2	318	13.0	0
v9	22	2	231	11.0	0
v9	23	2	265	12.0	0
v9	24	2	250	12.0	0
v9	25	2	257	12.0	0
v9	26	2	283	12.0	0
v9	27	2	197	11.0	0
v9	28	1	87	8.0	0
v9	29	2	231	10.0	0
v9	30	1	21	*	0
v9	31	1	20	*	0
v9	32	1	34	*	0
v9	33	1	35	*	0
v9	34	1	17	*	0
v9	35	1	20	*	0
v9	36	1	34	*	0
v10	1	1	108	9.0	0
v10	2	2	327	13.0	2
v10	3	2	475	14.0	2
v10	4	2	371	13.0	2
v10	5	2	406	14.0	2
v10	6	2	416	14.0	2
v10	7	2	544	16.0	2
v10	8	2	440	14.0	2
v10	9	2	350	13.0	2
v10	10	2	430	14.0	2
v10	11	2	410	14.0	2
v10	12	2	349	13.0	2
v10	13	2	36	8.0	2
v10	14	1	104	8.0	0
v10	15	1	113	7.0	0
v10	16	1	92	6.0	0
v10	17	1	35	*	0

APPENDIX 5.3

Temperature recordings at each site. Not all enclosures had a max-min thermometer. These enclosures are given the temperature of the nearest enclosure with a thermometer. Temperatures were recorded on five occasions at puddle enclosures and twice at pond enclosures.

Site	enc	Observed Max	min	Mid	Assigned to enclosure
PUDDLES					
2133		30.0	15.0		no enclosures survived
		23.0	14.0		at this site.
		24.0	15.0		
		26.0	16.0		
		25.0	16.0		
mean		**21.6**	**15.2**	**18.4**	
2126	P1	20.5	13.0		P1, P2,P 3, P4
		15.0	15.0		
		18.0	11.5		
		19.0	12.0		
		20.0	14.0		
mean		**28.5**	**13.1**	**15.8**	
2134		26.5	14.0		no enclosures survived
		22.0	14.0		at this site.
		22.5	14.0		
		23.0	15.0		
		23.0	16.0		
mean		**234**	**14.6**	**19.0**	
2140		*	13.0		P6
		16.0	13.0		This enclosure is in
		16.0	13.0		site 2139 where no
		17.0	14.0		max-min was
		17.5	14.5		placed. 2140 is the
mean		**16.6**	**13.5**	**15.1**	nearest site to it on same disused railway track.
PONDS					
Veleševec	V1	28.0	14.0	29.0	V1, V2, max-min in
		30.0	19.0		V1; V2 is
mean		**39.0**	**23.5**	**26.2**	adjacent
Veleševec	V3	26.5	18.0		V3, V4 max-min in
		31.0	20.0		V3; V4 is
mean		**38.5**	**19.0**	**23.7**	adjacent
1039	V5	28.0	14.0	28.0	V5, V6, V7 "
		28.0	18.0		
mean		**28.0**	**18.0**	**22.0**	
Vratova	V9	23.0	12.0	22.5	V8, V9, V10
		22.0	13.0		
mean		**22.5**	**12.5**	**17.5**	